

MINIMIZATION OF DIVERGENCES ON SETS OF SIGNED MEASURES

MICHEL BRONIATOWSKI* AND AMOR KEZIOU**

ABSTRACT. We consider the minimization problem of ϕ -divergences between a given probability measure P and subsets Ω of the vector space $\mathcal{M}_{\mathcal{F}}$ of all signed finite measures which integrate a given class \mathcal{F} of bounded or unbounded measurable functions. The vector space $\mathcal{M}_{\mathcal{F}}$ is endowed with the weak topology induced by the class $\mathcal{F} \cup \mathcal{B}_b$ where \mathcal{B}_b is the class of all bounded measurable functions. We treat the problems of existence and characterization of the ϕ -projections of P on Ω . We consider also the dual equality and the dual attainment problems when Ω is defined by linear constraints.

Key words: Minimum Divergences; Maximum Entropy; Convex Programming; Moment Problem; Empirical Likelihood; Convex Distances; Fenchel Duality.

MSC (2000) Classification: 49A55; 49A40; 46A20; 46A05; 62E20.

CONTENTS

1. Introduction and notation	1
2. Existence of ϕ -Projections on Sets of signed measures	8
3. Characterization of ϕ -Projections on sets of signed measures	16
3.1. On general Sets Ω	16
3.2. On Sets defined by Linear Constraints	18
4. Fenchel duality for ϕ -Divergences	21
5. Applications to the minimization of ϕ -divergences on sets of signed finite measures satisfying linear constraints	25
References	27

1. INTRODUCTION AND NOTATION

Let $(\mathcal{X}, \mathcal{B})$ be a measurable space and P be a given reference probability measure (p.m.) on $(\mathcal{X}, \mathcal{B})$. Denote \mathcal{M} the real vector space of all signed finite measures on $(\mathcal{X}, \mathcal{B})$ and $\mathcal{M}(P)$ the vector subspace of all signed finite measures absolutely continuous (a.c) with respect to (w.r.t.) P . Denote also \mathcal{M}^1 the set of all p.m.'s on $(\mathcal{X}, \mathcal{B})$ and $\mathcal{M}^1(P)$ the subset of all p.m.'s a.c w.r.t. P . Let φ be a proper¹ closed² convex function from $]-\infty, +\infty[$ to $[0, +\infty]$ with $\varphi(1) = 0$ and such that its domain $\text{dom} \varphi := \{x \in \mathbb{R} \text{ such that } \varphi(x) < \infty\}$ is an interval with endpoints $a_\varphi < 1 < b_\varphi$ (which

Date: March 2010.

THIS IS AN AN ELECTRONIC REPRINT OF THE ORIGINAL ARTICLE PUBLISHED IN STUDIA SCI. MATH. HUNGAR., 2006, VOL. 43, NO. 4, 403–442. THIS REPRINT DIFFERS FROM THE ORIGINAL IN PAGINATION AND TYPOGRAPHIC DETAIL.

¹We say a function is proper if its domain is non void.

²The closedness of φ means that if a_φ or b_φ are finite numbers then $\varphi(x)$ tends to $\varphi(a_\varphi)$ or $\varphi(b_\varphi)$ when $x \downarrow a_\varphi$ or $x \uparrow b_\varphi$, respectively.

may be finite or infinite). For any signed finite measure Q in $\mathcal{M}(P)$, the ϕ -divergence between Q and P is defined by

$$(1.1) \quad \phi(Q, P) := \int_{\mathcal{X}} \varphi \left(\frac{dQ}{dP}(x) \right) dP(x).$$

When Q is not a.c. w.r.t. P , we set $\phi(Q, P) = +\infty$. The ϕ -divergences between p.m.'s were introduced by Csiszár (1963) as “ f -divergences”. The definition of ϕ -divergences of Csiszár (1963) between p.m.'s requires a common dominating σ -finite measure, noted λ , for Q and P . Note that the two definitions of ϕ -divergences coincide on the set of all p.m.'s a.c. w.r.t. P and dominated by λ . The ϕ -divergences between any signed finite measure Q and a p.m. P were introduced by Csiszár *et al.* (1999); they gave the following definition

$$(1.2) \quad \phi(Q, P) := \int \varphi(q) dP + b\sigma_Q^+(\mathcal{X}) - a\sigma_Q^-(\mathcal{X}),$$

where $a := \lim_{x \rightarrow -\infty} \varphi(x)/x$, $b := \lim_{x \rightarrow +\infty} \varphi(x)/x$ and

$$Q = qP + \sigma_Q, \quad \sigma_Q = \sigma_Q^+ - \sigma_Q^-$$

is the Lebesgue decomposition of Q , and the Jordan decomposition of the singular part σ_Q , respectively. The definitions (1.1) and (1.2) coincide when Q is a.c. w.r.t. P or when $a = -\infty$ or $b = +\infty$. Since we will consider optimization of $Q \mapsto \phi(Q, P)$ on sets of signed finite measures a.c. w.r.t. P , it is more adequate for our sake to use the definition (1.1).

For all p.m. P , the mappings $Q \in \mathcal{M} \mapsto \phi(Q, P)$ are convex and take nonnegative values. When $Q = P$ then $\phi(Q, P) = 0$. Furthermore, if the function $x \mapsto \varphi(x)$ is strictly convex on a neighborhood of $x = 1$, then the following basic property holds

$$(1.3) \quad \phi(Q, P) = 0 \text{ if and only if } Q = P.$$

All these properties are presented in Csiszár (1963), Csiszár (1967a), Csiszár (1967b) and Liese and Vajda (1987) chapter 1, for ϕ -divergences defined on the set of all p.m.'s \mathcal{M}^1 . When the ϕ -divergences are defined on \mathcal{M} , then the same properties hold.

When defined on \mathcal{M}^1 , the Kullback-Leibler (KL), modified Kullback-Leibler (KL_m), χ^2 , modified χ^2 (χ_m^2), Hellinger (H), and L_1 divergences are respectively associated to the convex functions $\varphi(x) = x \log x - x + 1$, $\varphi(x) = -\log x + x - 1$, $\varphi(x) = \frac{1}{2}(x-1)^2$, $\varphi(x) = \frac{1}{2}(x-1)^2/x$, $\varphi(x) = 2(\sqrt{x}-1)^2$ and $\varphi(x) = |x-1|$. All those divergences except the L_1 one, belong to the class of power divergences introduced in Cressie and Read (1984) (see also Liese and Vajda (1987) chapter 2). They are defined through the class of convex functions

$$(1.4) \quad x \in]0, +\infty[\mapsto \varphi_\gamma(x) := \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma - 1)}$$

if $\gamma \in \mathbb{R} \setminus \{0, 1\}$, $\varphi_0(x) := -\log x + x - 1$ and $\varphi_1(x) := x \log x - x + 1$. (For all $\gamma \in \mathbb{R}$, we define $\varphi_\gamma(0) := \lim_{x \downarrow 0} \varphi_\gamma(x)$). So, the KL -divergence is associated to φ_1 , the KL_m to φ_0 , the χ^2 to φ_2 , the χ_m^2 to φ_{-1} and the Hellinger distance to $\varphi_{1/2}$.

The Kullback-Leibler divergence (KL -divergence) is sometimes called Boltzmann Shannon relative entropy. It appears in the domain of large deviations and it is frequently used for reconstruction of laws, and in particular in the classical moment problem (see e.g. Csiszár *et al.* (1999) and the references therein). The modified Kullback-Leibler divergence (KL_m -divergence) is sometimes called Burg relative entropy. It is frequently used in Statistics and it leads to efficient methods in statistical estimation and tests problems; in fact, the celebrate “maximum likelihood” method

can be seen as an optimization problem of the KL_m -divergence between the discrete or continuous parametric model and the empirical measure associated to the data; see Keziou (2003a) and Broniatowski and Keziou (2003). On the other hand, the recent “empirical likelihood” method can also be seen as an optimization problem of the KL_m -divergence between some set of measures satisfying some linear constraints and the empirical measure associated to the data; see Owen (2001) and the references therein, Bertail (2003), Bertail (2004) and Broniatowski and Keziou (2004). The Hellinger divergence is also used in Statistics, it leads to robust statistical methods in parametric and semi-parametric models; see Beran (1977), Lindsay (1994), Jiménez and Shao (2001) and Broniatowski and Keziou (2004).

We extend the definition of the power divergences functions $Q \in \mathcal{M}^1 \mapsto \phi_\gamma(Q, P)$ onto the whole vector space of signed finite measures \mathcal{M} via the extension of the definition of the convex functions φ_γ : For all $\gamma \in \mathbb{R}$ such that the function $x \mapsto \varphi_\gamma(x)$ is not defined on $] -\infty, 0[$ or defined but not convex on whole \mathbb{R} , we extend its definition as follows

$$(1.5) \quad x \in] -\infty, +\infty[\mapsto \begin{cases} \varphi_\gamma(x) & \text{if } x \in [0, +\infty[, \\ +\infty & \text{if } x \in] -\infty, 0[. \end{cases}$$

Note that for the χ^2 -divergence for instance, $\varphi_2(x) := \frac{1}{2}(x-1)^2$ is defined and convex on whole \mathbb{R} .

The conjugate (or Fenchel-Legendre transform) of φ will be denoted φ^* , i.e.,

$$(1.6) \quad t \in \mathbb{R} \mapsto \varphi^*(t) := \sup_{x \in \mathbb{R}} \{tx - \varphi(x)\},$$

and the endpoints of $\text{dom} \varphi^*$ (the domain of φ^*) will be denoted a_{φ^*} and b_{φ^*} with $a_{\varphi^*} \leq b_{\varphi^*}$. Note that φ^* is proper closed convex function. In particular, $a_{\varphi^*} < 0 < b_{\varphi^*}$, $\varphi^*(0) = 0$ and

$$(1.7) \quad a_{\varphi^*} = \lim_{y \rightarrow -\infty} \frac{\varphi(y)}{y}, \quad b_{\varphi^*} = \lim_{y \rightarrow +\infty} \frac{\varphi(y)}{y}.$$

By the closedness of φ , the conjugate φ^{**} of φ^* coincides with φ , i.e.,

$$(1.8) \quad \varphi^{**}(t) := \sup_{x \in \mathbb{R}} \{tx - \varphi^*(x)\} = \varphi(t), \quad \text{for all } t \in \mathbb{R}.$$

For the proper convex functions defined on \mathbb{R} (endowed with the usual topology), the lower semi-continuity³ and the closedness properties are equivalent.

We say that φ (resp. φ^*) is differentiable if it is differentiable on $]a_\varphi, b_\varphi[$ (resp. $]a_{\varphi^*}, b_{\varphi^*}[$), the interior of its domain. We say also that φ (resp. φ^*) is strictly convex if it is strictly convex on $]a_\varphi, b_\varphi[$ (resp. $]a_{\varphi^*}, b_{\varphi^*}[$).

The strict convexity of φ is equivalent to the condition that its conjugate φ^* is essentially smooth, i.e., differentiable with

$$(1.9) \quad \begin{aligned} \lim_{t \downarrow a_{\varphi^*}} \varphi^{*'}(t) &= -\infty & \text{if } a_{\varphi^*} > -\infty, \\ \lim_{t \uparrow b_{\varphi^*}} \varphi^{*'}(t) &= +\infty & \text{if } b_{\varphi^*} < +\infty. \end{aligned}$$

Conversely, φ is essentially smooth if and only if φ^* is strictly convex; see e.g. Rockafellar (1970) section 26 for the proofs of these properties.

If φ is differentiable, we denote φ' the derivative function of φ , and we define $\varphi'(a_\varphi)$ and $\varphi'(b_\varphi)$ to be the limits (which may be finite or infinite) $\lim_{x \downarrow a_\varphi} \varphi'(x)$ and $\lim_{x \uparrow b_\varphi} \varphi'(x)$, respectively. We denote $\text{Im} \varphi'$ the set of all values of the function φ' , i.e., $\text{Im} \varphi' := \{\varphi'(x) \text{ such that } x \in [a_\varphi, b_\varphi]\}$.

³We say a function φ is lower semi-continuous if the level sets $\{x \text{ such that } \varphi(x) \leq \alpha\}$, $\alpha \in \mathbb{R}$ are closed.

If additionally the function φ is strictly convex, then φ' is increasing on $[a_\varphi, b_\varphi]$. Hence, it is one-to-one function from $[a_\varphi, b_\varphi]$ to $\text{Im}\varphi'$, we denote in this case φ'^{-1} the inverse function of φ' from $\text{Im}\varphi'$ to $[a_\varphi, b_\varphi]$.

Note that if φ is differentiable, then for all $x \in]a_\varphi, b_\varphi[$,

$$(1.10) \quad \varphi^*(\varphi'(x)) = x\varphi'(x) - \varphi(x).$$

If additionally φ is strictly convex, then for all $t \in \text{Im}\varphi'$ we have

$$(1.11) \quad \varphi^*(t) = t\varphi'^{-1}(t) - \varphi(\varphi'^{-1}(t)) \quad \text{and} \quad \varphi^{*'}(t) = \varphi'^{-1}(t).$$

On the other hand, if φ is essentially smooth, then the interior of the domain of φ^* coincides with that of $\text{Im}\varphi'$, i.e., $(a_{\varphi^*}, b_{\varphi^*}) = (\varphi'(a_\varphi), \varphi'(b_\varphi))$.

The domain of the ϕ -divergence will be denoted $\text{dom}\phi$, i.e.,

$$(1.12) \quad \text{dom}\phi := \{Q \in \mathcal{M} \text{ such that } \phi(Q, P) < \infty\}.$$

Definition 1.1. Let Ω be some subset in \mathcal{M} . The ϕ -divergence between the set Ω and a p.m. P , noted $\phi(\Omega, P)$, is

$$\phi(\Omega, P) := \inf_{Q \in \Omega} \phi(Q, P).$$

Definition 1.2. Assume that $\phi(\Omega, P)$ is finite. A measure $Q^* \in \Omega$ such that

$$\phi(Q^*, P) \leq \phi(Q, P) \quad \text{for all } Q \in \Omega$$

is called a ϕ -projection of P on Ω . This projection may not exist, or may be not defined uniquely.

If φ is a strictly convex, then the function $Q \in \mathcal{M}(P) \mapsto \phi(Q, P)$ is strictly convex, and the ϕ -projection of P on some convex set Ω is uniquely defined whenever it exists.

Let $g_i : \mathcal{X} \mapsto \mathbb{R}$, $i = 1, \dots, l$, be measurable real valued functions on \mathcal{X} . Denote $g := (g_0, g_1, \dots, g_l)^T$ with $g_0 := \mathbf{1}_{\mathcal{X}}$. We assume that the functions g_0, g_1, \dots, g_l are linearly independent in the following sense : $P\{\lambda^T g(x) \neq 0\} > 0$ for any $\lambda \in \mathbb{R}^{1+l}$ with $\lambda \neq 0$. For all $\lambda \in \mathbb{R}^{1+l}$, we denote $\lambda_0, \lambda_1, \dots, \lambda_l$ the $(1+l)$ coordinates of λ .

Let's denote by M_g the set of all signed finite measures with total mass one, a.c. w.r.t. P , which integrate the functions g_i and satisfy a finite number of linear constraints, i.e.,

$$(1.13) \quad M_g := \left\{ Q \in \mathcal{M}(P) \text{ such that } Q(\mathcal{X}) = 1 \text{ and } \int_{\mathcal{X}} g_i(x) dQ(x) = 0, i = 1, \dots, l \right\}.$$

We consider the optimization problem

$$(1.14) \quad \inf_{Q \in M_g} \phi(Q, P).$$

The Lagrangian “dual” problem associated with (1.14) is

$$(1.15) \quad \sup_{\lambda \in \mathbb{R}^{1+l}} \left\{ \lambda_0 - \int_{\mathcal{X}} \varphi^*(\lambda^T g(x)) dP(x) \right\}.$$

We will consider the problem of the “dual” equality $\inf(1.14) = \sup(1.15)$, the existence of optimal solutions in (1.15), and in particular the problems of the existence and the characterization of the optimal solutions in (1.14), i.e., the ϕ -projections of P on the set M_g .

These problems intervene in the domain of the reconstruction of laws, in particular, the classical moment problem. Also they appear frequently in Statistics; in fact, the recent “empirical likelihood” method, which is the non parametric version of the celebrated maximum likelihood method, can be seen as an optimization problem of the KL_m -divergence between some set of measures defined as in (1.13) and the empirical measure associated to the data.

In the vocabulary of the duality theory, a measure Q in M_g which realizes the infimum in (1.14) (i.e., a ϕ -projection of P on M_g in the vocabulary of ϕ -divergences theory) is called “a primal optimal solution” or simply “an optimal solution”, while a point λ in \mathbb{R}^{1+l} realizing the supremum in (1.15) is called “a dual optimal solution”.

For the optimization problem of convex function $\psi : \mathbb{R}^n \mapsto]-\infty, +\infty]$ on convex sets C in \mathbb{R}^n subject to linear constraints $Ax = b \in \mathbb{R}^m$ where A is some $m \times n$ -matrix, a sufficient condition, in order that the equality

$$(1.16) \quad \inf_{\{x \in C; Ax=b\}} \psi(x) = \sup_{t \in \mathbb{R}^m} \{b^T t - \psi^*(A^T t)\}$$

holds with dual attainment, is that there exists a point \hat{x} in the relative interior⁴ of the convex set $C \cap \text{dom} \psi$ such that $A\hat{x} = b$. See e.g. Rockafellar (1970) for the proofs of these results.

In order to make the set M_g closed and the linear functions $Q \in \mathcal{M} \mapsto \int_{\mathcal{X}} g_i(x) dQ(x)$ continuous (which we need to apply the duality theory and to treat the problem of existence of ϕ -projections of P on the set M_g), we endow the vector space \mathcal{M} by the weak topology which we denote $\tau_{\mathcal{F}}$ induced by $\mathcal{F} \cup \mathcal{B}_b$ where $\mathcal{F} := \{g_0, g_1, \dots, g_l\}$ and \mathcal{B}_b is the set of all bounded \mathcal{B} -measurably real valued functions on \mathcal{X} ; see section 2 below for precise definition of the $\tau_{\mathcal{F}}$ -topology.

Note that the relative interior of the convex set M_g is generally empty in the weak topology $\tau_{\mathcal{F}}$. Borwein and Lewis (1992) have extended the idea of the relative interior (r.i.) of convex sets in \mathbb{R}^n to a new notion which have called “the quasi relative interior” (q.r.i.) of convex subsets of an arbitrary Hausdorff topological vector space X (having finite or infinite dimension), and they used it to construct a powerful duality theory for the optimization problem of convex function $\psi : X \mapsto (-\infty, +\infty]$ on convex sets $C \subseteq X$ subject to linear constraints. In particular, when X is locally convex, they obtain similar results as in (1.16) when the relative interior is replaced by the quasi relative interior; see Borwein and Lewis (1992) Corollary 4.8. The main advantage of the quasi relative interior of convex subset C of infinite dimension vector space X is that it is frequently nonempty even when the relative interior of C is empty.

If $\int_{\mathcal{X}} |g_i(x)| dP(x)$ is finite for all $i = 1, \dots, l$, then the convex conjugate of the convex function $Q \mapsto \phi(Q, P)$ (on the vector space $\mathcal{M}_{\mathcal{F}}(P)$ of all signed finite measures Q a.c. w.r.t. P and which integrate all the elements of \mathcal{F} , i.e., all the functions g_i) can be written as

$$\phi^*(f) := \sup_{Q \in \mathcal{M}_{\mathcal{F}}(P)} \left\{ \int f dQ - \phi(Q, P) \right\} = \int \varphi^*(f) dP, \text{ for all } f \in \langle \mathcal{F} \cup \mathcal{B}_b \rangle;$$

see section 4 below for details. So, in this case, as in Borwein and Lewis (1991), we can apply Corollary 4.8 of Borwein and Lewis (1992) to obtain the dual equality $\inf(1.14) = \sup(1.15)$ with dual attainment, whenever there exists a measure Q_0 in M_g which belongs to the quasi relative interior of $M_g \cap \text{dom} \phi$. This condition is called “constraint qualification”. We can prove also from Borwein and Lewis (1992) that a measure Q_0 is in the q.r.i. of $M_g \cap \text{dom} \phi$ iff $a_{\varphi} < \frac{dQ_0}{dP} \leq \frac{dQ_0}{dP} < b_{\varphi}$,

⁴i.e., the interior in the real affine subspace $\langle C \cap \text{dom} \psi \rangle$ of \mathbb{R}^n endowed with the relative topology of the usual topology on \mathbb{R}^n .

P -almost everywhere (P -a.e.). We summarize these results and some other results about the problems of the existence and the characterization of the primal optimal solution (i.e., the ϕ -projection of P on M_g) in the following two Theorems and two Corollaries. For proofs, see Theorem 3.10 of Borwein and Lewis (1992), Corollary 2.6 and Theorem 4.8 of Borwein and Lewis (1991), and Theorem II.2 of Csiszár *et al.* (1999).

Theorem 1.1. *If $\int_{\mathcal{X}} |g_i(x)| dP(x)$ is finite for all $i = 1, \dots, l$, and if the following constraint qualification⁵:*

$$(1.17) \quad \text{there is a } \widehat{Q} \in M_g \cap \text{dom}\phi \text{ such that } a_\varphi < \frac{d\widehat{Q}}{dP} \leq \frac{d\widehat{Q}}{dP} < b_\varphi \quad (P - \text{a.e.})$$

holds, then $\inf(1.14) = \sup(1.15)$ and there is attainment in (1.15). Suppose additionally that φ^ is essentially smooth (which is equivalent to the strict convexity of φ), and that there exists a dual optimal solution $\bar{\lambda}$ which is an interior point of*

$$(1.18) \quad \text{dom}\phi^* := \left\{ \lambda \in \mathbb{R}^{1+l} \text{ such that } \int_{\mathcal{X}} \varphi^*(\lambda^T g(x)) dP(x) \text{ is finite} \right\}.$$

Then the unique optimal solution of (1.14) (i.e., the ϕ -projection of P on M_g), which we denote by Q^ , exists and it is given by*

$$(1.19) \quad \frac{dQ^*}{dP}(x) = \varphi^{*'}(\bar{\lambda}^T g(x)).$$

In (1.18), for brevity, the definition of $\text{dom}\phi^*$, which usually is the set of functions f such that $\phi^*(f) < \infty$, is modified here.

Remark 1.1. *If all functions g_i belong to $L_\infty(\mathcal{X}, P)$, and if for a dual optimal solution $\bar{\lambda} \in \mathbb{R}^{1+l}$ the following condition*

$$(1.20) \quad a_{\varphi^*} < \text{essinf } \bar{\lambda}^T g(\cdot) \leq \text{esssup } \bar{\lambda}^T g(\cdot) < b_{\varphi^*}$$

holds, then $\bar{\lambda}$ is an interior point of $\text{dom}\phi^$. Hence, under assumption (1.20), all results in the above Theorem hold whenever the constraint qualification (1.17) is met.*

If all functions g_i belong to $L_\infty(\mathcal{X}, P)$, and the convex function φ^* is everywhere finite (i.e., $a_{\varphi^*} = -\infty$ and $b_{\varphi^*} = +\infty$), then obviously condition (1.20) holds since $\text{dom}\phi^* = \mathbb{R}^{1+l}$ in this case. Hence, under the constraint qualification (1.17), all results in the above Theorem hold. We state this result in the following Corollary.

Corollary 1.2. *Suppose that all functions g_i belong to $L_\infty(\mathcal{X}, P)$ and φ^* is everywhere finite (i.e., $a_{\varphi^*} = -\infty$ and $b_{\varphi^*} = +\infty$). If the constraint qualification (1.17) holds, then $\inf(1.14) = \sup(1.15)$ and there is attainment in (1.15). Suppose additionally that φ^* is everywhere differentiable (which is equivalent to the strict convexity of φ), then the unique optimal solution Q^* of (1.14) (i.e., the ϕ -projection of P on M_g) exists and it is given by*

$$(1.21) \quad \frac{dQ^*}{dP}(x) = \varphi^{*'}(\bar{\lambda}^T g(x)),$$

where $\bar{\lambda} \in \mathbb{R}^{1+l}$ is any dual optimal solution.

⁵The strict inequalities in (1.17) mean that $P \left\{ \frac{d\widehat{Q}}{dP} \leq a_\varphi \right\} = P \left\{ \frac{d\widehat{Q}}{dP} \geq b_\varphi \right\} = 0$.

In the following Theorem and Corollary, we give sufficient conditions for the uniqueness of the dual optimal solution (see Borwein and Lewis (1991) Theorem 4.5 for the proof). Note that the strict convexity of φ^* is equivalent to the condition that its conjugate φ is essentially smooth.

Theorem 1.3. *Suppose that all assumptions of Theorem 1.1 are satisfied. Suppose furthermore that the function φ is essentially smooth. Then the dual optimal solution $\bar{\lambda}$ is unique. Moreover, the unique optimal solution Q^* of (1.14) exists and it is given by*

$$(1.22) \quad \frac{dQ^*}{dP}(x) = \varphi^{*'}(\bar{\lambda}^T g(x)) = \varphi'^{-1}(\bar{\lambda}^T g(x)).$$

Corollary 1.4. *Suppose that all assumptions of Corollary 1.2 are satisfied. Suppose additionally that the function φ is essentially smooth. Then the dual optimal solution $\bar{\lambda}$ is unique. Moreover, the unique optimal solution Q^* of (1.14) exists and it is given by*

$$(1.23) \quad \frac{dQ^*}{dP}(x) = \varphi^{*'}(\bar{\lambda}^T g(x)) = \varphi'^{-1}(\bar{\lambda}^T g(x)).$$

The important Corollary 1.2, which essentially requires that the constraint qualification (1.17) holds, applies in the KL -divergence case since the corresponding conjugate φ^* is everywhere finite (see also Borwein and Lewis (1993) for other examples), but it fails in the two important cases of Burg relative entropy (KL_m -divergence in the context of divergences) and Hellinger divergence without additional conditions since the corresponding conjugates φ^* are infinite on the intervals $[1, +\infty)$ and $[2, +\infty)$, respectively.

Léonard (2001b) consider the optimization problem (1.14) when the set M_g is replaced by the subset

$$(1.24) \quad M_{os} := \left\{ Q \in \mathcal{M}(P) \text{ such that } q := \frac{dQ}{dP} \in L_{\varphi_m^{**}}, \int_{\mathcal{X}} g(x) dQ = (1, 0, \dots, 0)^T \right\},$$

where $L_{\varphi_m^{**}}$ is the Orlicz space defined as follows:

$$(1.25) \quad L_{\varphi_m^{**}} := \{ q : \mathcal{X} \rightarrow \mathbb{R}; \text{ measurable such that } \|q\|_{\varphi_m^{**}} < \infty \}$$

with $\|q\|_{\varphi_m^{**}} := \inf \left\{ a > 0; \int_{\mathcal{X}} \varphi_m^{**} \left(\frac{|q(x)|}{a} \right) dP(x) \leq 1 \right\},$

and φ_m^{**} is the convex conjugate of the convex function φ_m^* defined by $\varphi_m^*(t) := \max(\varphi^*(t), \varphi^*(-t))$ for all $t \in \mathbb{R}$. Without the constraint qualification (1.17), under the following integrability condition

$$(1.26) \quad \text{for any } \lambda \in \mathbb{R}^{1+l}, \int_{\mathcal{X}} \varphi^*(\lambda^T g(x)) dP(x) < \infty,$$

applying the duality theory on Orlicz spaces, Léonard (2001b) obtains the dual equality

$$(1.27) \quad \inf_{Q \in M_{os}} \phi(Q, P) = \sup_{\lambda \in \mathbb{R}^{1+l}} \left\{ \lambda_0 - \int_{\mathcal{X}} \varphi^*(\lambda^T g(x)) dP(x) \right\}.$$

Moreover, if the value is finite, then there exists at least one ϕ -projection of P on M_{os} ; see Theorem 3.4 of Léonard (2001b) for details in more general context. A characterization of the ϕ -projections of P on the set M_{os} (with finite or infinite number of linear constraints) is stated by Léonard (2001c) under condition (1.26); see Theorems 4.4, 4.5 and 4.6 of Léonard (2001c). Note that the integrability condition (1.26) implies that φ^* is everywhere finite, and these results apply in the important KL -divergence case with finite or infinite number of linear constraints. However, the condition (1.26) does not hold in the KL_m -divergence and Hellinger divergence cases since the

domains of the corresponding φ^* functions are proper subsets of \mathbb{R} , and the important result (1.27) does not apply in these two important cases. Under the weaker integrability assumption

$$(1.28) \quad \text{for any } \lambda \in \mathbb{R}^{1+l}, \text{ there exists } \alpha > 0 \text{ such that} \\ \int_{\mathcal{X}} \varphi^*(\alpha \lambda^T g(x)) dP(x) + \int_{\mathcal{X}} \varphi^*(-\alpha \lambda^T g(x)) dP(x) < \infty,$$

the dual equality (1.27) may fail; see Theorem 3.3 of Léonard (2001b).

The goal of the present paper is to give results of existence and characterization of the ϕ -projections of a given p.m. P on some subsets Ω of $\mathcal{M}_{\mathcal{F}}$, the space of all signed finite measures which integrate a given class \mathcal{F} of functions, in particular, convex sets of signed finite measures defined by linear constraints as in (1.13) extending some previous works (about the existence and characterization of the ϕ -projections on subsets of \mathcal{M}^1 , the set of all p.m.'s) of Csiszár (1975), Liese (1977), Csiszár (1984), Rüschemdorf (1984), Rüschemdorf (1987), Liese and Vajda (1987), Teboulle and Vajda (1993) and Csiszár (1995). We give also different versions of dual representations of the ϕ -divergences viewed as convex functions on the space of all signed finite measures which integrate an arbitrary class of functions. When the set Ω is defined by linear constraints as in (1.13), we consider the dual problem, and we obtain the equality $\inf(1.14) = \sup(1.15)$ with dual attainment under different assumptions without constraint qualification. Additional conditions are given to obtain similar results which apply in the two important KL_m -divergence and Hellinger divergence cases.

Enhancing \mathcal{M}^1 to \mathcal{M} is motivated by the following arguments: sometimes the ϕ -projection, say Q_1^* , of a p.m. P on subset of \mathcal{M}^1 is not an “interior” point and we can not give in this case a definite description of Q_1^* , while the ϕ -projection, say Q^* , of a p.m. P on subset of \mathcal{M} is an “interior” point, which allows to give a perfect characterization of the ϕ -projection Q^* (see example 3.1). In the context of statistical estimation and tests using the empirical likelihood method (see Owen (2001)), or related ones to criterions defined through divergences (see Broniatowski and Keziou (2004)), the projection of the empirical measure P_n of a sample on a set Ω^1 of p.m.'s may make problems when the projection is not an interior point of $\Omega^1 \cap \text{dom}\phi(\cdot, P_n)$. Enhancing \mathcal{M}^1 to \mathcal{M} , this difficulty does not hold any longer, and tests as well as estimation can be performed.

The rest of this paper is organized as follows : In section 2, we consider the problem of existence of ϕ -projections on general closed sets of signed measures. In section 3, we deal with the problem of characterization of the ϕ -projections on sets of signed measures, in particular, sets of signed measures defined by linear constraints. In section 4, we give different dual representations of ϕ -divergences seen as convex functions on the vector space of all signed finite measures which integrate a given class of functions. In section 5, we apply the results of sections 2, 3 and 4, to obtain the dual equality $\inf(1.14) = \sup(1.15)$ with dual attainment, under different assumptions without constraint qualification.

2. EXISTENCE OF ϕ -PROJECTIONS ON SETS OF SIGNED MEASURES

In this section, we give sufficient conditions for the existence of ϕ -projections of some p.m. P on sets Ω of signed finite measures which integrate some class of functions (see Theorems 2.5, 2.6 and 2.7, and Corollary 2.8 below). At first, we give some notation and we establish a convenient topological context for this problem. Let \mathcal{F} be some class of measurable real valued functions f (bounded or unbounded) defined on \mathcal{X} . Here, \mathcal{F} is not assumed to be finite. Denote by \mathcal{B}_b the set of all bounded measurable real valued functions defined on \mathcal{X} , and by $\langle \mathcal{F} \cup \mathcal{B}_b \rangle$ the linear span of

$\mathcal{F} \cup \mathcal{B}_b$. Define the set

$$\mathcal{M}_{\mathcal{F}}^1 := \left\{ Q \in \mathcal{M}^1 \text{ such that } \int |f| dQ < \infty, \text{ for all } f \text{ in } \mathcal{F} \right\},$$

and the real vector space

$$\mathcal{M}_{\mathcal{F}} := \left\{ Q \in \mathcal{M} \text{ such that } \int |f| d|Q| < \infty, \text{ for all } f \text{ in } \mathcal{F} \right\},$$

in which $|Q|$ denotes the total variation of the signed finite measure Q .

Note that if $\mathcal{F} = \mathcal{B}_b$, then $\mathcal{M}_{\mathcal{F}}^1 = \mathcal{M}^1$ and $\mathcal{M}_{\mathcal{F}} = \mathcal{M}$.

Definition 2.1. Denote by $\tau_{\mathcal{F}}$ the weakest topology on $\mathcal{M}_{\mathcal{F}}$ for which all mappings $Q \in \mathcal{M}_{\mathcal{F}} \mapsto \int f dQ$ are continuous when f belongs to $\mathcal{F} \cup \mathcal{B}_b$. Denote also by $\tau_{\mathcal{M}}$ the weakest topology on $\langle \mathcal{F} \cup \mathcal{B}_b \rangle$ for which all mappings $f \in \langle \mathcal{F} \cup \mathcal{B}_b \rangle \mapsto \int f dQ$ are continuous when $Q \in \mathcal{M}_{\mathcal{F}}$. We sometimes call $\tau_{\mathcal{F}}$ the topology induced by $\langle \mathcal{F} \cup \mathcal{B}_b \rangle$ on $\mathcal{M}_{\mathcal{F}}$, and likewise $\tau_{\mathcal{M}}$ the topology induced by $\mathcal{M}_{\mathcal{F}}$ on $\langle \mathcal{F} \cup \mathcal{B}_b \rangle$.

A base of open neighborhoods for any R in $\mathcal{M}_{\mathcal{F}}$ is defined by

$$(2.1) \quad U(R, \mathcal{A}, \varepsilon) := \left\{ Q \in \mathcal{M}_{\mathcal{F}} \text{ such that } \max_{f \in \mathcal{A}} \left| \int f dR - \int f dQ \right| < \varepsilon \right\}$$

for $\varepsilon > 0$ and \mathcal{A} a finite collection of functions in $\langle \mathcal{F} \cup \mathcal{B}_b \rangle$.

We refer to Chapter 5 of Dunford and Schwartz (1962), for the various topologies induced by classes of functions. Note that the class \mathcal{B}_b induces the so-called τ -topology (see e.g. Groeneboom *et al.* (1979) and Gänssler (1971)), and that $\mathcal{M}_{\mathcal{B}_b}$ is the whole vector space \mathcal{M} .

The above $\tau_{\mathcal{F}}$ -topology on $\mathcal{M}_{\mathcal{F}}$ is indeed the natural and the most convenient one in order to handle projection properties. It has been introduced in the context of large deviation probabilities by Eichelsbacher and Schmock (2002) for the Kullback-Leibler divergence and it is used in Statistics in Broniatowski (2003), Keziou (2003a), Broniatowski and Keziou (2003) and Keziou (2003b). Usually the sets which are to be considered in statistical applications are not compact but merely closed sets; a typical example is when they are defined by linear constraints as in (1.13). Hence, the set M_g is closed in $\mathcal{M}_{\mathcal{F}}$ endowed with the $\tau_{\mathcal{F}}$ -topology if the functions g_i (which may be bounded or unbounded) belong to \mathcal{F} ; this motivates the choice of $\tau_{\mathcal{F}}$ -topology.

Proposition 2.1. Equip $\mathcal{M}_{\mathcal{F}}$ with the $\tau_{\mathcal{F}}$ -topology and $\langle \mathcal{F} \cup \mathcal{B}_b \rangle$ with the $\tau_{\mathcal{M}}$ -topology. Then, $\mathcal{M}_{\mathcal{F}}$ and $\langle \mathcal{F} \cup \mathcal{B}_b \rangle$ are Hausdorff locally convex topological vector spaces. Further, the topological dual space of $\mathcal{M}_{\mathcal{F}}$ is the set of all mappings $Q \mapsto \int f dQ$ when f belongs to $\langle \mathcal{F} \cup \mathcal{B}_b \rangle$, and the topological dual space of $\langle \mathcal{F} \cup \mathcal{B}_b \rangle$ is the set of all mappings $f \mapsto \int f dQ$ when Q belongs to $\mathcal{M}_{\mathcal{F}}$.

Proof of Proposition 2.1 By Lemma 5.3.3 in Dunford and Schwartz (1962), the vector space $\mathcal{M}_{\mathcal{F}}$ equipped with the $\tau_{\mathcal{F}}$ -topology is a Hausdorff locally convex topological space. On the other hand, the set of all mappings $Q \in \mathcal{M}_{\mathcal{F}} \mapsto \int f dQ$ when f belongs to $\langle \mathcal{F} \cup \mathcal{B}_b \rangle$ is a total linear space; indeed, for all $Q \in \mathcal{M}_{\mathcal{F}}$, assume that $\int f dQ = 0$ for all f in $\langle \mathcal{F} \cup \mathcal{B}_b \rangle$, choose $f = \mathbb{1}_{\{B\}}$ for any $B \in \mathcal{B}$ to conclude that $Q = 0$. The proof ends then as a consequence of Theorem 5.3.9 in Dunford and Schwartz (1962). ■

We denote by $[\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}]$ and by $[\langle \mathcal{F} \cup \mathcal{B}_b \rangle; \tau_{\mathcal{M}}]$ the two Hausdorff locally convex topological vector spaces endowed with the $\tau_{\mathcal{F}}$ -topology and the $\tau_{\mathcal{M}}$ -topology, respectively.

Broniatowski and Keziou (2003) have proved that the function $Q \in [\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}] \mapsto \phi(Q, P)$ is lower semi-continuous (l.s.c.), provided only that the corresponding convex function φ is closed; see Proposition 2.3 of Broniatowski and Keziou (2003) and Proposition 2.1 of Keziou (2003a) which we recall here for convenience.

Proposition 2.2. *For any ϕ -divergence, the divergence function $Q \mapsto \phi(Q, P)$ from $[\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}]$ to $[0, +\infty]$ is l.s.c.*

We will use the following Lemma to prove Proposition 2.2.

Lemma 2.3. *Let $\mathcal{M}_{\mathcal{F}}(P)$ denotes the vector subspace of all signed measures in $\mathcal{M}_{\mathcal{F}}$ which are absolutely continuous w.r.t. P . The vector subspace $\mathcal{M}_{\mathcal{F}}(P)$ is a closed set in $[\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}]$.*

Proof of Lemma 2.3 Let $\overline{\mathcal{M}_{\mathcal{F}}(P)}$ denotes the closure of $\mathcal{M}_{\mathcal{F}}(P)$ in $[\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}]$. Assume that there exists R in $\overline{\mathcal{M}_{\mathcal{F}}(P)}$ with R not in $\mathcal{M}_{\mathcal{F}}(P)$. Then, there exists some B in \mathcal{B} such that $P(B) = 0$ and $R(B) \neq 0$. On the other hand, for all n in \mathbb{N} , the set $U := U(R, \mathbb{1}_{\{B\}}, 1/n)$ is a neighborhood of R (see (2.1)), hence, $U \cap \mathcal{M}_{\mathcal{F}}(P)$ is non void. Therefore, we can construct a sequence of measures R_n in $\mathcal{M}_{\mathcal{F}}(P)$ such that

$$\left| \int \mathbb{1}_{\{B\}} dR - \int \mathbb{1}_{\{B\}} dR_n \right| < 1/n.$$

Since $R_n(B) = 0$ for all n in \mathbb{N} , we deduce that $R(B) = 0$, a contradiction. This implies that $\overline{\mathcal{M}_{\mathcal{F}}(P)} = \mathcal{M}_{\mathcal{F}}(P)$, that is $\mathcal{M}_{\mathcal{F}}(P)$ is closed in $[\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}]$. This concludes the proof of Lemma 2.3. \blacksquare

Remark 2.1. *Note that if $\mathcal{F} = \mathcal{B}_b$, then $\mathcal{M}_{\mathcal{F}} = \mathcal{M}$ and $\mathcal{M}_{\mathcal{F}}(P) = \mathcal{M}(P)$. Hence, we deduce from Lemma 2.3 that the subspace $\mathcal{M}(P)$ is closed in $[\mathcal{M}; \tau]$, the space of all signed finite measures endowed with the τ -topology. Note also that $\mathcal{M}_{\mathcal{F}}^1$ and $\mathcal{M}_{\mathcal{F}}^1(P)$ are closed in $[\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}]$, and that \mathcal{M}^1 and $\mathcal{M}^1(P)$ are closed in $[\mathcal{M}; \tau]$.*

Proof of Proposition 2.2 Let α be a real number. We prove that the set

$$A(\alpha) := \{Q \in \mathcal{M}_{\mathcal{F}} \text{ such that } \phi(Q, P) \leq \alpha\}$$

is closed in $[\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}]$. By Lemma 2.3, $\mathcal{M}_{\mathcal{F}}(P)$ is closed in $[\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}]$. Since $A(\alpha)$ is included in $[\mathcal{M}_{\mathcal{F}}(P); \tau_{\mathcal{F}}]$, we have to prove that $A(\alpha)$ is closed in the subspace $[\mathcal{M}_{\mathcal{F}}(P); \tau_{\mathcal{F}}]$. Let

$$B(\alpha) := \left\{ f \in L_1(\mathcal{X}, P) \text{ such that } \int \varphi(f(x)) dP(x) \leq \alpha \right\}.$$

$B(\alpha)$ is a convex set, since φ is a convex function. Furthermore, $B(\alpha)$ is closed in $L_1(\mathcal{X}, P)$. Indeed, let f_n be a sequence in $B(\alpha)$ with $\lim_{n \rightarrow \infty} f_n = f^*$, where the limit is intended in $L_1(\mathcal{X}, P)$. Hence, there exists a subsequence f_{n_k} which converges to f^* (P -a.e.). The functions $\varphi(f_{n_k})$ are nonnegative. Further, we have $\liminf_{k \rightarrow +\infty} \varphi(f_{n_k}(x)) = \varphi(f^*(x))$ (P -a.e.) by the closedness of the convex function φ . Therefore, Fatou's Lemma implies

$$\int \varphi(f^*) dP \leq \liminf_{k \rightarrow +\infty} \int \varphi(f_{n_k}) dP \leq \liminf_{k \rightarrow +\infty} \int \varphi(f_{n_k}) dP \leq \alpha,$$

which is to say that f^* belongs to $B(\alpha)$. Hence, $B(\alpha)$ is a closed subset in $L_1(\mathcal{X}, P)$. Since, it is convex, it is then weakly closed in $L_1(\mathcal{X}, P)$; see e.g. Theorem 5.3.13 in Dunford and Schwartz

(1962). Denote by W the weak topology on $L_1(\mathcal{X}, P)$ and consider the mapping H defined by

$$\begin{aligned} H &: [\mathcal{M}_{\mathcal{F}}(P); \tau_{\mathcal{F}}] \mapsto [L_1(\mathcal{X}, P); W] \\ Q &\mapsto H(Q) = dQ/dP. \end{aligned}$$

Let us prove that H is weakly continuous, that is $Q \mapsto \int H(Q)g \, dP$ is a continuous mapping for all g in $L_{\infty}(\mathcal{X}, P)$. Indeed, let g be some function in $L_{\infty}(\mathcal{X}, P)$. Then, we have

$$\int H(Q)g \, dP = \int (dQ/dP)g \, dP = \int g \, dQ.$$

The mapping $Q \mapsto \int g \, dQ$ is $\tau_{\mathcal{F}}$ -continuous; indeed, for all g in $L_{\infty}(\mathcal{X}, P)$, it holds $P(g > \|g\|_{\infty}) = 0$, which implies $Q(g > \|g\|_{\infty}) = 0$, for all Q in $\mathcal{M}_{\mathcal{F}}(P)$. Therefore, $\int g \, dQ = \int g \mathbf{1}_{[g \leq \|g\|_{\infty}]} \, dQ$. Now, the mapping $Q \mapsto \int g \mathbf{1}_{[g \leq \|g\|_{\infty}]} \, dQ$ is continuous in $\tau_{\mathcal{F}}$ -topology since $g \mathbf{1}_{[g \leq \|g\|_{\infty}]} \in \mathcal{F} \cup \mathcal{B}_b$. Since $A(\alpha) = \{Q \in \mathcal{M}_{\mathcal{F}}(P), \phi(Q, P) \leq \alpha\} = H^{-1}(B(\alpha))$, we deduce that $A(\alpha)$ is closed in $[\mathcal{M}_{\mathcal{F}}(P); \tau_{\mathcal{F}}]$, for any α in \mathbb{R} . This proves Proposition 2.2. ■

For any ϕ -divergence, by the lower semi-continuity of the function $Q \in [\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}] \mapsto \phi(Q, P)$, the following result holds.

Theorem 2.4. *Let P be some p.m. and Ω some compact subset of $[\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}]$. Then there exists at least one ϕ -projection of P on Ω .*

Using some similar arguments as used in the proof of Theorem 2.4 in Liese (1977) or Proposition 8.5 in Liese and Vajda (1987) and Fenchel's inequality or Hölder inequality, we state general results for the existence of ϕ -projections of some p.m. P on closed sets Ω of $[\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}]$ (see Theorem 2.6 and 2.7 below). At first, in the following Theorem, we give a version of Theorem 2.4 in Liese (1977) or Proposition 8.5 in Liese and Vajda (1987).

Theorem 2.5. *Let Ω be some closed set in $[\mathcal{M}; \tau]$. Assume that the following two conditions*

$$(2.2) \quad \phi(\Omega, P) := \inf_{Q \in \Omega} \phi(Q, P) < \infty^6$$

and

$$(2.3) \quad \lim_{|x| \rightarrow \infty} \frac{\varphi(x)}{|x|} = +\infty$$

hold. Then there exists at least one ϕ -projection of P on Ω .

Proof of Theorem 2.5 Denote $m := \phi(\Omega, P)$ which is finite by assumption, and let β be a positive number. Define the sets

$$\Omega(\beta) := \{Q \in \Omega \text{ such that } \phi(Q, P) \leq m + \beta\}$$

and

$$\Lambda(\beta) := \left\{ q := \frac{dQ}{dP} \text{ such that } Q \in \Omega(\beta) \right\}.$$

The set $\Lambda(\beta)$ is uniformly integrable. Hence, it is weakly sequentially compact in $L_1(\mathcal{X}, P)$, (see e.g. Meyer (1966) p. 39). Consider now a sequence Q_n in $\Omega(\beta)$ such that

$$\lim_{n \rightarrow +\infty} \phi(Q_n, P) = \phi(\Omega, P).$$

⁶Note that this is equivalent to the condition: there exists $Q \in \Omega$ such that $\phi(Q, P)$ is finite.

The sequence $q_n := dQ_n/dP$ belongs to $\Lambda(\beta)$. Therefore, there exists a subsequence $(q_{n_i})_{i \in \mathbb{N}}$ which converges weakly in $L_1(\mathcal{X}, P)$ to some function $q^* \in L_1(\mathcal{X}, P)$, which is to say that the corresponding sequence of signed finite measures Q_{n_i} converges to $Q^* \in \mathcal{M}(P)$ in τ -topology where Q^* is defined by $dQ^*/dP := q^*$. Hence, Q^* belongs to Ω since it is the limit in τ -topology of the sequence (Q_{n_i}) which belongs to the τ -closed set Ω . On the other hand, the mapping $Q \in [\mathcal{M}; \tau] \mapsto \phi(Q, P)$ is l.s.c.⁷, and therefore

$$(2.4) \quad \phi(Q^*, P) \leq \lim_{i \rightarrow +\infty} \phi(Q_{n_i}, P) = \phi(\Omega, P) < \infty.$$

We deduce that Q^* is a ϕ -projection of P on Ω . ■

Remark 2.2. For sets Ω of p.m.'s defined by linear constraints, sufficient conditions for the existence of KL -projections are presented in (Csiszár (1975) Theorem 3.1, Corollary 3.1 and Theorem 3.3). Sufficient conditions of the existence of ϕ -projections on sets of p.m.'s satisfying linear equality or inequality constraints are given in Csiszár (1995) Theorem 3.

Remark 2.3. By Eichelsbacher and Schmock (2002), if for all $\alpha > 0$ and all $f \in \mathcal{F}$, $\int \exp(\alpha|f|) dP < \infty$, then the level sets

$$\{Q \in \mathcal{M}_{\mathcal{F}}^1 \text{ such that } KL(Q, P) \leq c\}$$

are compact in $[\mathcal{M}_{\mathcal{F}}^1; \tau_{\mathcal{F}}]$ for all real c . Therefore, for any $\tau_{\mathcal{F}}$ -closed set $\Omega \subset \mathcal{M}_{\mathcal{F}}^1$ for which $KL(\Omega, P) < \infty$, the projection of P on Ω exists; see Eichelsbacher and Schmock (2002) Lemma 2.1.

Using Fenchel's inequality and some similar arguments to that in Lemma 2.1 of Eichelsbacher and Schmock (2002), We generalize Theorem 3 of Csiszár (1995) and the result in Remark 2.3 about the existence of projections, to the class of ϕ -divergences and to $\tau_{\mathcal{F}}$ -closed sets of signed measures.

Theorem 2.6. Let Ω be some closed set in $\mathcal{M}_{\mathcal{F}}$ equipped with the $\tau_{\mathcal{F}}$ -topology. Suppose that the following three assumptions

$$(2.5) \quad \phi(\Omega, P) < \infty,$$

$$(2.6) \quad \lim_{|x| \rightarrow \infty} \frac{\varphi(x)}{|x|} = +\infty$$

$$(2.7) \quad \text{and for every } f \in \mathcal{F} \text{ and every } \alpha > 0, \int \varphi^*(\alpha|f|) dP < \infty$$

hold. Then there exists at least one ϕ -projection of P on Ω .

Proof of Theorem 2.6 As shown in the proof of Theorem 2.5, under assumptions (2.5) and (2.6), there exists a sequence $(Q_{n_i})_{i \in \mathbb{N}}$ in $\Omega(\beta) \subset \Omega$ that converges in τ -topology to some Q^* in $\mathcal{M}(P)$ satisfying

$$(2.8) \quad \phi(Q^*, P) \leq \lim_{i \rightarrow +\infty} \phi(Q_{n_i}, P) = \phi(\Omega, P) < \infty.$$

⁷this holds from Proposition 2.2 choosing the class of functions $\mathcal{F} = \mathcal{B}_b$, the class of all bounded measurable real valued functions.

It remains to prove that Q^* belongs to Ω . At first, we prove that Q^* belongs to $\mathcal{M}_{\mathcal{F}}$. So, let f in \mathcal{F} . Denote by Q_+^* the nonnegative variation and by Q_-^* the nonpositive variation of Q^* : $Q^* = Q_+^* - Q_-^*$. Using Fenchel's inequality through the integral we can write

$$\begin{aligned}
 \int |f| dQ_+^* &= \int |f| q_+^* dP \\
 &\leq \int \varphi(q_+^*) dP + \int \varphi^*(|f|) dP \\
 &\leq \int \varphi(q^*) dP + \int \varphi^*(|f|) dP \\
 (2.9) \quad &= \phi(Q^*, P) + \int \varphi^*(|f|) dP,
 \end{aligned}$$

and similarly

$$(2.10) \quad \int |f| dQ_-^* \leq \phi(Q^*, P) + \int \varphi^*(|f|) dP.$$

Hence, from (2.8), (2.9) and (2.10), we deduce $\int |f| d|Q^*| < \infty$ since

$$\int |f| d|Q^*| = \int |f| dQ_+^* + \int |f| dQ_-^*.$$

Hence Q^* belongs to $\mathcal{M}_{\mathcal{F}}$. We still have to prove that Q^* belongs to Ω . Since Ω is, by assumption, a closed set in $[\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}]$, it is enough to show that the sequence $(Q_{n_i})_i$ (which belongs to $\Omega(\beta) \subset \Omega$) converges to Q^* in $[\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}]$. Note that the sequence Q_{n_i} converges to Q^* in τ -topology. Hence, we still have to prove that $\int f dQ_{n_i}$ converges to $\int f dQ^*$ for all f in \mathcal{F} . So, let f in \mathcal{F} . We use now similar argument as in the proof of Lemma 2.1 in Eichelsbacher and Schmock (2002). Let $\epsilon > 0$. Define $\alpha = (m + \beta)/\epsilon$. Using the fact that $\varphi^*(0) = 0$, by condition (2.7) and the dominated convergence theorem, there exists $j_0 \in \mathbb{N}$ such that

$$\frac{1}{\alpha} \int \varphi^*(\alpha |f| \mathbb{1}_{\{|f| > j\}}) dP < \epsilon$$

for all $j \geq j_0$. Hence, using Fenchel's inequality and the fact that the sequence $(Q_{n_i})_i$ belongs to $\Omega(\beta)$, we can write

$$\begin{aligned}
 \left| \int f dQ_{n_i} - \int f \mathbb{1}_{\{|f| \leq j\}} dQ_{n_i} \right| &\leq \int |f - f \mathbb{1}_{\{|f| \leq j\}}| d|Q_{n_i}| \\
 &= \frac{1}{\alpha} \int \alpha |f| \mathbb{1}_{\{|f| > j\}} d|Q_{n_i}| \\
 &\leq 2 \left[\frac{1}{\alpha} \phi(Q_{n_i}, P) + \frac{1}{\alpha} \int \varphi^*(\alpha |f| \mathbb{1}_{\{|f| > j\}}) dP \right] \\
 (2.11) \quad &\leq 2 \left[\frac{1}{\alpha} (m + \beta) + \epsilon \right] = 4\epsilon.
 \end{aligned}$$

We have just proved that, for all $f \in \mathcal{F}$, for all $\epsilon > 0$, there exists $j_0 \in \mathbb{N}$, such that for all $j \geq j_0$ and all $i \in \mathbb{N}$,

$$(2.12) \quad \int f \mathbb{1}_{\{|f| \leq j\}} dQ_{n_i} - 4\epsilon \leq \int f dQ_{n_i} \leq \int f \mathbb{1}_{\{|f| \leq j\}} dQ_{n_i} + 4\epsilon.$$

Using the fact that the sequence $(Q_{n_i})_i$ converges to Q^* in τ -topology, by passage to limits in (2.12) when $i \rightarrow \infty$, then when $j \rightarrow \infty$ and finally when $\epsilon \rightarrow 0$, we get $\lim_{i \rightarrow \infty} \int f dQ_{n_i} = \int f dQ^*$. Hence, the sequence $(Q_{n_i})_i$ converges to Q^* in $\tau_{\mathcal{F}}$ -topology, which implies that Q^* belongs to Ω since Ω is closed in $[\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}]$. From the inequality (2.8), we conclude that Q^* is a ϕ -projection of

P on Ω . This completes the proof. ■

Using Hölder inequality, we give in the following Theorem another result of existence of ϕ -projection on closed set in $[\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}]$. In the sequel, $\|\cdot\|_k$ denotes the usual norm of the vector space $L_k(\mathcal{X}, P)$, $1 \leq k \leq +\infty$.

Theorem 2.7. *Let Ω be some closed set in $\mathcal{M}_{\mathcal{F}}$ equipped with the $\tau_{\mathcal{F}}$ -topology. Assume that the following conditions*

$$(2.13) \quad \phi(\Omega, P) < \infty,$$

$$(2.14) \quad \text{there exists numbers } 1 < r, k < +\infty \text{ such that } r^{-1} + k^{-1} = 1,$$

$$\lim_{|x| \rightarrow \infty} \frac{\varphi(x)}{|x|^r} > 0, \quad \text{and for every } f \in \mathcal{F}, \quad \|f\|_k < \infty$$

hold. Then there exists at least one ϕ -projection of P on Ω .

Proof of Theorem 2.7 Since condition (2.14) implies (2.6), as in the proof of Theorem 2.5, there exists a sequence $(Q_{n_i})_{i \in \mathbb{N}}$ in $\Omega(\beta) \subset \Omega$ that converges in τ -topology to some Q^* in $\mathcal{M}(P)$ satisfying

$$(2.15) \quad \phi(Q^*, P) \leq \lim_{i \rightarrow +\infty} \phi(Q_{n_i}, P) = \phi(\Omega, P) < \infty.$$

We have to prove that Q^* belongs to Ω . At first, we prove that Q^* belongs to $\mathcal{M}_{\mathcal{F}}$. For all f in \mathcal{F} , we have

$$\begin{aligned} \int |f| \, d|Q^*| &= \int |f| |q^*| \, dP \\ &= \int |f| |q^*| \mathbf{1}_{\{|q^*| \leq c_0\}} \, dP + \int |f| |q^*| \mathbf{1}_{\{|q^*| > c_0\}} \, dP \\ &\leq c_0 \int |f| \, dP + \int |f| \frac{|q^*|}{\varphi(q^*)^{1/r}} \varphi(q^*)^{1/r} \mathbf{1}_{\{|q^*| > c_0\}} \, dP \\ &\leq c_0 \int |f| \, dP + c_1 \left(\int |f|^k \, dP \right)^{1/k} \left(\int \varphi(q^*) \, dP \right)^{1/r} \\ (2.16) \quad &= c_0 \int |f| \, dP + c_1 \left(\int |f|^k \, dP \right)^{1/k} (\phi(Q^*, P))^{1/r}. \end{aligned}$$

Hence, from (2.15) and (2.14), we deduce $\int |f| \, d|Q^*| < \infty$. We still have to prove that Q^* belongs to Ω . Since Ω is, by assumption, a closed set in $[\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}]$, it is enough to show that the sequence $(Q_{n_i})_i$ (which belongs to $\Omega(\beta) \subset \Omega$) converges to Q^* in $[\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}]$. Note that the sequence Q_{n_i} converges to Q^* in τ -topology. Hence, we still have to prove that $\int f \, dQ_{n_i}$ converges to $\int f \, dQ^*$ for all f in \mathcal{F} . So, let f in \mathcal{F} . For all positive number b , using (2.14), we can write

$\int f dQ_{n_i} = \int f \mathbb{1}_{\{|f| \leq b\}} dQ_{n_i} + \int f \mathbb{1}_{\{|f| > b\}} dQ_{n_i} =: A + B$, and

$$\begin{aligned}
|B| &= \left| \int f \mathbb{1}_{\{|f| > b\}} dQ_{n_i} \right| \leq \int |f| \mathbb{1}_{\{|f| > b\}} d|Q_{n_i}| = \int |f| \mathbb{1}_{\{|f| > b\}} |q_{n_i}| dP \\
&= \int |f| \mathbb{1}_{\{|f| > b\}} |q_{n_i}| \mathbb{1}_{\{|q_{n_i}| \leq c_0\}} dP + \int |f| \mathbb{1}_{\{|f| > b\}} |q_{n_i}| \mathbb{1}_{\{|q_{n_i}| > c_0\}} dP \\
&\leq c_0 \int |f| \mathbb{1}_{\{|f| > b\}} dP + \int |f| \mathbb{1}_{\{|f| > b\}} \frac{|q_{n_i}|}{\varphi(q_{n_i})^{1/r}} \varphi(q_{n_i})^{1/r} \mathbb{1}_{\{|q_{n_i}| > c_0\}} dP \\
&\leq c_0 \int |f| \mathbb{1}_{\{|f| > b\}} dP + c_1 \int |f| \mathbb{1}_{\{|f| > b\}} \varphi(q_{n_i})^{1/r} dP \\
&\leq c_0 \int |f| \mathbb{1}_{\{|f| > b\}} dP + c_1 \left(\int |f|^k \mathbb{1}_{\{|f| > b\}} dP \right)^{1/k} \left(\int \varphi(q_{n_i}) dP \right)^{1/r}.
\end{aligned}$$

We deduce

$$(2.17) \quad (B1) \leq \int f dQ_{n_i} \leq (B2),$$

with

$$(B1) := \int f \mathbb{1}_{\{|f| \leq b\}} dQ_{n_i} - c_0 \int |f| \mathbb{1}_{\{|f| > b\}} dP - c_1 \left(\int |f|^k \mathbb{1}_{\{|f| > b\}} dP \right)^{1/k} \left(\int \varphi(q_{n_i}) dP \right)^{1/r},$$

and

$$(B2) := \int f \mathbb{1}_{\{|f| \leq b\}} dQ_{n_i} + c_0 \int |f| \mathbb{1}_{\{|f| > b\}} dP + c_1 \left(\int |f|^k \mathbb{1}_{\{|f| > b\}} dP \right)^{1/k} \left(\int \varphi(q_{n_i}) dP \right)^{1/r}.$$

The functions $\{f_b := |f| \mathbb{1}_{\{|f| > b\}}, b \geq 0\}$ and $\{f_b^k := |f|^k \mathbb{1}_{\{|f| > b\}}, b \geq 0\}$ are dominated respectively by $|f|$ and $|f|^k$. Moreover, $\int |f| dP$ and $\int |f|^k dP$ are finite by assumption (2.14). We thus get by the dominated convergence theorem

$$\lim_{b \rightarrow +\infty} \int |f| \mathbb{1}_{\{|f| > b\}} dP = \lim_{b \rightarrow +\infty} \int |f|^k \mathbb{1}_{\{|f| > b\}} dP = 0.$$

Hence, from (2.17), we get

$$\int f dQ^* = \lim_{b \rightarrow +\infty} \lim_{i \rightarrow +\infty} (B1) \leq \lim_{i \rightarrow +\infty} \int f dQ_{n_i} \leq \lim_{b \rightarrow +\infty} \lim_{i \rightarrow +\infty} (B2) = \int f dQ^*,$$

which is to say that the subsequence $(Q_{n_i})_i$ converges to Q^* in $\tau_{\mathcal{F}}$ -topology. Hence, Q^* belongs to Ω . From inequality (2.15), we deduce that Q^* is a ϕ -projection of P on Ω . This ends the proof of Theorem 2.7. ■

Note that the above results do not apply in the case of KL_m and Hellinger divergences since the condition $\lim_{|x| \rightarrow \infty} \frac{\varphi(x)}{|x|} = +\infty$ does not hold. The following Corollary applies without assumption $\lim_{|x| \rightarrow \infty} \frac{\varphi(x)}{|x|} = +\infty$, in particular, in the KL_m and Hellinger divergences cases.

Corollary 2.8. *Let Ω be a closed set in $[\mathcal{M}, \tau]$. If the following condition: there exists*

$$(2.18) \quad u, l \in L_1(\mathcal{X}, P) \text{ such that } u \leq \frac{dQ}{dP} \leq l \text{ (} P\text{-a.e.) for all } Q \in \Omega \cap \text{dom}\phi$$

holds, then there exists at least one ϕ -projection of P on Ω whenever $\phi(\Omega, P)$ is finite.

Proof of Corollary 2.8 Similar to that of Theorem 2.5. The uniform integrability of the set $\Lambda(\beta)$ holds by condition (2.18). ■

3. CHARACTERIZATION OF ϕ -PROJECTIONS ON SETS OF SIGNED MEASURES

In this section, we extend known results pertaining to the characterization of the ϕ -projections as can be found in Rüschendorf (1984), Rüschendorf (1987), Liese and Vajda (1987), (see also Csiszár (1975) and Csiszár (1984) for the characterization of KL -projections). These authors have characterized the ϕ -projections on subsets of \mathcal{M}^1 . We expose similar results when considering subsets of \mathcal{M} and take the occasion to clarify some proofs. We first consider the case of general subsets Ω of \mathcal{M} and then the case of convex subsets of \mathcal{M} defined by linear constraints. For the whole Section, we assume that the convex function φ is differentiable.

3.1. On general Sets Ω . We will use the following assumption

$$(3.1) \quad \begin{aligned} &\text{There exists } 0 < \delta < 1 \text{ such that for all } c \text{ in } [1 - \delta, 1 + \delta], \\ &\text{we can find numbers } c_1, c_2, c_3 \text{ such that} \\ &\varphi(cx) \leq c_1\varphi(x) + c_2|x| + c_3, \text{ for all real } x. \end{aligned}$$

Remark 3.1. Condition (3.1) holds for all power divergences including KL , KL_m and Hellinger divergences. Note also that condition (3.1) implies that a_φ equals 0 or $-\infty$ and b_φ equals $+\infty$.

Remark 3.2. In all the sequel, condition (3.1) above can be replaced by any other condition which implies part (1) of Lemma 3.1 below.

We first give two Lemmas, which we will use in the proof of Theorem 3.3 and Theorem 3.4 below.

Lemma 3.1. Assume that (3.1) holds. Then, for all Q in \mathcal{M} such that $\phi(Q, P)$ is finite, we have

- (1) for any c in $[1 - \delta, 1 + \delta]$, $\varphi\left(c\frac{dQ}{dP}\right)$ belongs to $L_1(\mathcal{X}, P)$.
- (2) $\lim_{c \uparrow 1} \phi(cQ, P) = \phi(Q, P) = \lim_{c \downarrow 1} \phi(cQ, P)$.

Proof of Lemma 3.1 (1) Under condition (3.1), for all Q in \mathcal{M} such that $\phi(Q, P) < \infty$, we have

$$\varphi\left(c\frac{dQ}{dP}\right) \leq c_1\varphi\left(\frac{dQ}{dP}\right) + c_2\left|\frac{dQ}{dP}\right| + c_3.$$

Integrating with respect to P yields

$$\int \varphi\left(c\frac{dQ}{dP}\right) dP \leq c_1\phi(Q, P) + c_2 \int \left|\frac{dQ}{dP}\right| dP + c_3 < \infty.$$

(2) For all c in $[1 - \delta, 1 + \delta]$, define the functions

$$\begin{aligned} l_c &: x \in \mathbb{R} \mapsto l_c(x) := \varphi(cx)\mathbf{1}_{]-\infty, 0[}(cx), \\ g_c &: x \in \mathbb{R} \mapsto g_c(x) := \varphi(cx)\mathbf{1}_{[0, 1]}(cx), \\ h_c &: x \in \mathbb{R} \mapsto h_c(x) := \varphi(cx)\mathbf{1}_{]1, +\infty[}(cx). \end{aligned}$$

For any c and x , we have $\varphi(cx) = l_c(x) + g_c(x) + h_c(x)$. For all real x , the functions $c \rightarrow l_c(x)$ and $c \rightarrow h_c(x)$ are nondecreasing, and the function $c \rightarrow g_c(x)$ is nonincreasing. Denote $q := \frac{dQ}{dP}$. Apply the monotone convergence theorem to get

$$\lim_{c \uparrow 1} \int l_c(q) dP = \int l_1(q) dP \quad \text{and} \quad \lim_{c \uparrow 1} \int h_c(q) dP = \int h_1(q) dP.$$

On the other hand, the class of functions $\{x \rightarrow g_c(x), c \text{ in } [1 - \delta, 1 + \delta]\}$ is bounded above by the function $x \rightarrow g_{1-\delta}(x)$. Furthermore, for all Q in \mathcal{M} , $g_{1-\delta}(q)$ belongs to $L_1(\mathcal{X}, P)$ by the condition

(3.1). Hence, applying the monotone convergence theorem we get

$$\lim_{c \uparrow 1} \int g_c(q) dP = \int g_1(q) dP.$$

Those three limits prove the first part of the claim. The same argument completes the proof of the Lemma. ■

Lemma 3.2. *Assume that condition (3.1) holds. Then, for all Q in $\text{dom}\phi$, $\varphi'(q)q$ belongs to $L_1(\mathcal{X}, P)$, where $q := \frac{dQ}{dP}$.*

Proof of Lemma 3.2 Using the convexity of the function φ , for all $\epsilon > 0$, we have

$$\frac{\varphi(q) - \varphi((1-\epsilon)q)}{\epsilon} \leq q\varphi'(q) \leq \frac{\varphi((1+\epsilon)q) - \varphi(q)}{\epsilon}.$$

By Lemma 3.1, for all ϵ satisfying $0 < \epsilon < \delta$, both the LHS and the RHS terms belong to $L_1(\mathcal{X}, P)$, and hence $\varphi'(q)q \in L_1(\mathcal{X}, P)$. ■

Theorem 3.3. *Let Ω be a subset of \mathcal{M} and Q^* be a signed measure in $\Omega \cap \text{dom}\phi$. Then*

- (1) *The following are sufficient conditions for Q^* to be a ϕ -projection of P on Ω : (i) $\varphi'(q^*)q \in L_1(\mathcal{X}, P)$ and (ii) $\int \varphi'(q^*) dQ^* \leq \int \varphi'(q^*) dQ$, for all Q in $\Omega \cap \text{dom}\phi$.*
- (2) *If condition (3.1) holds and Ω is convex, then these conditions are necessary as well.*

Proof of Theorem 3.3 Convexity and differentiability of φ imply, for all positive ϵ ,

$$(3.2) \quad \varphi'(q^*)(q - q^*) \leq \frac{\varphi((1-\epsilon)q^* + \epsilon q) - \varphi(q^*)}{\epsilon} \leq \varphi(q) - \varphi(q^*).$$

The middle term in the above display, by the convexity of φ , decreases to $\varphi'(q^*)(q - q^*)$ when $\epsilon \downarrow 0$. Furthermore, it is bounded above by $\varphi(q) - \varphi(q^*)$ which belongs to $L_1(\mathcal{X}, P)$ for all Q in $\text{dom}\phi$. Hence, applying the monotone convergence theorem to get

$$(3.3) \quad \int \varphi'(q^*)(q - q^*) dP = \lim_{\epsilon \downarrow 0} \int \frac{\varphi((1-\epsilon)q^* + \epsilon q) - \varphi(q^*)}{\epsilon} dP, \quad \text{for all } Q \in \text{dom}\phi.$$

Proof of part (1): Integrating (3.2) with respect to P and using (i) and (ii) in part (1) of the Theorem, we obtain for all Q in $\Omega \cap \text{dom}\phi$

$$(3.4) \quad \phi(Q, P) - \phi(Q^*, P) \geq \int \varphi'(q^*)(q - q^*) dP = \int \varphi'(q^*) dQ - \int \varphi'(q^*) dQ^* \geq 0.$$

Hence, Q^* is a ϕ -projection of P on Ω . Proof of part (2): Convexity of both Ω and $\text{dom}\phi$, implies that for all $Q \in \Omega \cap \text{dom}\phi$, $(1-\epsilon)Q + \epsilon Q^*$ belongs to $\Omega \cap \text{dom}\phi$. Since Q^* is a ϕ -projection of P on Ω , for all $Q \in \Omega \cap \text{dom}\phi$ and all ϵ satisfying $0 < \epsilon < 1$, we get $\phi((1-\epsilon)Q + \epsilon Q^*, P) - \phi(Q^*, P) \geq 0$. Combining this with (3.3) and using the fact that Q^* is a ϕ -projection of P on Ω , we obtain for all Q in $\Omega \cap \text{dom}\phi$

$$(3.5) \quad \begin{aligned} \int \varphi'(q^*)(q - q^*) dP &= \lim_{\epsilon \downarrow 0} \int \frac{\varphi((1-\epsilon)q^* + \epsilon q) - \varphi(q^*)}{\epsilon} dP \\ &= \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} [\phi((1-\epsilon)Q^* + \epsilon Q, P) - \phi(Q^*, P)] \geq 0. \end{aligned}$$

On the other hand, integrating (3.2) with respect to P , we obtain for all Q in $\Omega \cap \text{dom}\phi$

$$(3.6) \quad \int \varphi'(q^*)(q - q^*) dP \leq \phi(Q, P) - \phi(Q^*, P) < \infty.$$

Hence, (3.5) and (3.6) imply

$$(3.7) \quad \varphi'(q^*)(q - q^*) \in L_1(\mathcal{X}, P), \text{ for all } Q \in \Omega \cap \text{dom}\phi.$$

By Lemma 3.2, $\varphi'(q^*)q^* \in L_1(\mathcal{X}, P)$. Combining this with (3.7), we obtain that

$$\text{for all } Q \in \Omega \cap \text{dom}\phi, \text{ we have } \varphi'(q^*)Q \in L_1(\mathcal{X}, P)$$

and $\int \varphi'(q^*) dQ^* \leq \int \varphi'(q^*) dQ$. This completes the proof of Theorem 3.3. ■

3.2. On Sets defined by Linear Constraints. In this subsection, we consider the problems of existence and characterization of ϕ -projections of some p.m. P on linear set S of measures in \mathcal{M} defined by arbitrary family of constraints. So, let \mathcal{G} denote a collection (finite or infinite, countable or not) of real valued functions defined on $(\mathcal{X}, \mathcal{B})$. The class \mathcal{G} is assumed to contain the function $\mathbb{1}_{\mathcal{X}}$. The set S is defined by

$$(3.8) \quad S := \left\{ Q \in \mathcal{M}_{\mathcal{G}}(P) \text{ such that } \int_{\mathcal{X}} dQ = 1, \int_{\mathcal{X}} g dQ = 0, \text{ for all } g \text{ in } \mathcal{G} \setminus \{\mathbb{1}_{\mathcal{X}}\} \right\}.$$

The following result states the explicit form of Q^* , a ϕ -projection of P on S , when it exists.

Theorem 3.4.

- (1) Let Q^* be some finite measure in $S \cap \text{dom}\phi$. A sufficient condition, for Q^* to be a ϕ -projection of P on S , is that there exists numbers $c_1, \dots, c_d \in \mathbb{R}$ and functions $g_1, \dots, g_d \in \mathcal{G}$ such that $\varphi'(q^*(x)) = c_1 g_1(x) + \dots + c_d g_d(x)$ (P -a.e.).
- (2) Assume that condition (3.1) holds. Then, any ϕ -projection, say Q^* , of P on S , if it exists, satisfies $\varphi'(q^*)$ belongs to $\overline{\langle \mathcal{G} \rangle}$, (the closure of $\langle \mathcal{G} \rangle$) in $L_1(\mathcal{X}, |Q^*|)$.

If \mathcal{G} is a finite collection of functions in $L_1(\mathcal{X}, |Q^*|)$, then the vector space $\langle \mathcal{G} \rangle$ is closed in $L_1(\mathcal{X}, |Q^*|)$. So, from the above Theorem, we can state the following Corollary:

Corollary 3.5. Let $\mathcal{G} := \{\mathbb{1}_{\mathcal{X}}, g_1, \dots, g_l\}$ be a finite collection of measurable functions on \mathcal{X} . Then (1) and (2) below hold.

- (1) Let Q^* be some measure in $S \cap \text{dom}\phi$. A sufficient condition, for Q^* to be a ϕ -projection of P on S , is that there exists some constant $c \in \mathbb{R}^{1+l}$ such that

$$\varphi' \left(\frac{dQ^*}{dP}(x) \right) = c_0 + \sum_{i=1}^l c_i g_i(x) \quad (P - \text{a.e.}).$$

- (2) Assume that condition (3.1) holds. Then any ϕ -projection, say Q^* , of P on S , if it exists, satisfies

there exists some constant $c \in \mathbb{R}^{1+l}$ such that

$$\varphi' \left(\frac{dQ^*}{dP}(x) \right) = c_0 + \sum_{i=1}^l c_i g_i(x) \quad (|Q^*| - \text{a.e.}).$$

It should be noticed that the preceding Theorem and Corollary do not provide a definite description of the projected measure; indeed, it does not give any information on the support of $|Q^*|$ (see example 3.1 below). However, if $\varphi(0) = +\infty$ (which holds for example for the KL_m -divergence), then any ϕ -projection Q^* of P on some set Ω , if it exists, has obviously the same support as P when $\phi(\Omega, P)$ is finite. Furthermore, we prove in the following Lemma that if $\varphi'(0) = -\infty$ (which holds for instance in the case of KL , KL_m and Hellinger divergences), then any ϕ -projection of

P on some convex set Ω when it exists has the same support as P . At first, state the following Corollary which applies in the KL_m -divergence case.

Corollary 3.6. *Let \mathcal{G} be defined as in Corollary 3.5. Assume that assumption (3.1) holds. Suppose additionally that $\varphi(0) = +\infty$, and let Q^* be some p.m. in $S \cap \text{dom}\phi$. Then Q^* is a ϕ -projection of P on S iff there exists some constant $c \in \mathbb{R}^{1+l}$ such that*

$$\varphi' \left(\frac{dQ^*}{dP}(x) \right) = c_0 + \sum_{i=1}^l c_i g_i(x) \quad (P - \text{a.e.}).$$

Lemma 3.7. *Assume that condition (3.1) holds, $a_\varphi = 0$ and $\varphi'(0) = -\infty$. Let Ω be some convex set of signed finite measures. If there exists some $Q_0 \in \Omega \cap \text{dom}\phi$ such that $\frac{dQ_0}{dP} > 0$ (P -a.e.), then any ϕ -projection, say Q^* , of P on Ω , if it exists, has the same support as P , i.e., $\frac{dQ^*}{dP} > 0$ (P -a.e.).*

Proof of Lemma 3.7 Let $A := \{x \in \mathcal{X}; q^*(x) = 0\}$. Suppose that $P(A) > 0$. Since Q_0 and P have the same support by assumption, $Q_0(A) > 0$. By (3.2) (replacing Q by Q_0), $Q_0(A) > 0$ implies that $\int \varphi'(q^*) q dP = -\infty$ since $\int |\varphi'(q^*) q^*| dP < \infty$. This contradicts (3.5), which completes the proof. ■

We can now state, from the above Theorem, the following Corollary which applies in the case of KL , KL_m and Hellinger divergences.

Corollary 3.8. *Let \mathcal{G} be defined as in Corollary 3.5. Assume that assumption (3.1) holds. Suppose additionally that $a_\varphi = 0$ and $\varphi'(0) = -\infty$. If there exists some $Q_0 \in S \cap \text{dom}\phi$ such that $\frac{dQ_0}{dP} > 0$ (P -a.e.), then the following holds : a p.m. Q^* in $S \cap \text{dom}\phi$ is a ϕ -projection of P on S iff there exists some constant $c \in \mathbb{R}^{1+l}$ such that*

$$\varphi' \left(\frac{dQ^*}{dP}(x) \right) = c_0 + \sum_{i=1}^l c_i g_i(x) \quad (P - \text{a.e.}).$$

Remark 3.3. *Versions of Theorem 3.4, for sets of p.m.'s, have been proved by Csiszár (1975) and Csiszár (1984) for the Kullback-Leibler divergence, and by Rüschemdorf (1984) and Liese and Vajda (1987) for ϕ -divergences between p.m.'s. We prove it in the present context, that is when the set S (see (3.8)) is a subset of signed finite measures and P is a p.m. using similar techniques.*

Proof of Theorem 3.4 We start by proving (1). If $\varphi'(q^*)$ belongs to $\langle \mathcal{G} \rangle$, then for all Q in S , we have $\int \varphi'(q^*) dQ^* = \int \varphi'(q^*) dQ$ which, by the first part of Theorem 3.3, proves that Q^* is a ϕ -projection of P on S . Proof of part (2): Since Q^* is a signed finite measure, by the Hahn decomposition theorem, there exists a partition $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$ such that $\mathcal{X}_1, \mathcal{X}_2 \in \mathcal{B}$ and satisfying

$$\text{for all } B \in \mathcal{B}, \text{ such that } B \subset \mathcal{X}_1 \text{ we have } Q^*(B) \geq 0$$

and

$$\text{for all } B \in \mathcal{B}, \text{ such that } B \subset \mathcal{X}_2 \text{ we have } Q^*(B) \leq 0.$$

Denote by Q_+^* and Q_-^* respectively the nonnegative variation and the nonpositive variation of Q^* which are defined, for all $B \in \mathcal{B}$, by

$$Q_+^*(B) := Q^*(B \cap \mathcal{X}_1) \quad \text{and} \quad Q_-^*(B) := -Q^*(B \cap \mathcal{X}_2).$$

So, Q_+^* and Q_-^* are nonnegative finite measures, $Q = Q_+^* - Q_-^*$ and the total variation $|Q^*|$ is, by definition, the nonnegative measure $Q_+^* + Q_-^*$. Denote by $\langle \mathcal{G} \rangle_+^\perp$ and by $\langle \mathcal{G} \rangle_-^\perp$ respectively the orthogonal of $\langle \mathcal{G} \rangle$ in $L_1(\mathcal{X}, Q_+^*)$ and in $L_1(\mathcal{X}, Q_-^*)$, i.e., the sets defined by

$$\langle \mathcal{G} \rangle_+^\perp := \left\{ h \in L_\infty(\mathcal{X}, Q_+^*) \text{ such that } \int f h dQ_+^* = 0, \text{ for all } f \in \langle \mathcal{G} \rangle \right\}$$

and

$$\langle \mathcal{G} \rangle_-^\perp := \left\{ h \in L_\infty(\mathcal{X}, Q_-^*) \text{ such that } \int f h dQ_-^* = 0, \text{ for all } f \in \langle \mathcal{G} \rangle \right\}.$$

We will prove that the two following assertions hold

$$(3.9) \quad \text{for all } h \in \langle \mathcal{G} \rangle_+^\perp, \text{ we have } \int \varphi'(q^*) h dQ_+^* = 0$$

and

$$(3.10) \quad \text{for all } h \in \langle \mathcal{G} \rangle_-^\perp, \text{ we have } \int \varphi'(q^*) h dQ_-^* = 0.$$

We prove (3.9) by deriving a contradiction: assume that there exists h in $\langle \mathcal{G} \rangle_+^\perp$ such that $\int \varphi'(q^*) h dQ_+^* \neq 0$. We then have either (a) $\int \varphi'(q^*) h dQ_+^* < 0$ or (b) $\int \varphi'(q^*) h dQ_+^* > 0$. Assume (a). For $0 < \epsilon < \delta$,⁸ define the measure Q_0 by

$$(3.11) \quad dQ_0 := \left(1 + \epsilon \frac{h \mathbb{1}_{\mathcal{X}_1}}{\|h\|_\infty} \right) dQ^*.$$

Then Q_0 belongs to S , and, following condition (3.1), Q_0 belongs to $\text{dom} \phi$ by Lemma 3.1. Furthermore,

$$\int \varphi'(q^*) dQ_0 = \int \varphi'(q^*) dQ^* + \epsilon \frac{1}{\|h\|_\infty} \int \varphi'(q^*) h dQ_+^* < \int \varphi'(q^*) dQ^*,$$

which contradicts the fact that Q^* is a ϕ -projection of P on S (see part 2 in Theorem 3.3). Assume (b). Consider $-h$ instead of h . We thus have proved (3.9). The same arguments hold for the proof of (3.10). Therefore, $\varphi'(q^*)$ belongs to $\left(\langle \mathcal{G} \rangle_+^\perp \right)_+^\perp$ and to $\left(\langle \mathcal{G} \rangle_-^\perp \right)_-^\perp$ respectively the orthogonal of $\langle \mathcal{G} \rangle_+^\perp$ in $L_1(\mathcal{X}, Q_+^*)$ and the orthogonal of $\langle \mathcal{G} \rangle_-^\perp$ in $L_1(\mathcal{X}, Q_-^*)$. By Hahn-Banach Theorem (see e.g. Section 2 of Brezis (1983)), we have

$$\left(\langle \mathcal{G} \rangle_+^\perp \right)_+^\perp = \overline{\langle \mathcal{G} \rangle}_+ \quad \text{and} \quad \left(\langle \mathcal{G} \rangle_-^\perp \right)_-^\perp = \overline{\langle \mathcal{G} \rangle}_-$$

which are respectively the closure of $\langle \mathcal{G} \rangle$ in $L_1(\mathcal{X}, Q_+^*)$ and the closure of $\langle \mathcal{G} \rangle$ in $L_1(\mathcal{X}, Q_-^*)$. This implies that $\varphi'(q^*) \in \overline{\langle \mathcal{G} \rangle}$ that is, $\varphi'(q^*)$ belongs to the closure of $\langle \mathcal{G} \rangle$ in $L_1(\mathcal{X}, |Q^*|)$. This completes the proof of Theorem 3.4. ■

Example 3.1. Let $\mathcal{X} := [0, 1]$, P be the uniform distribution on $[0, 1]$ and $\mathcal{G} := \{\mathbb{1}_{[0,1]}, I_d\}$ where I_d is the identity function. Consider the χ_+^2 -divergence associated to the convex function

$$\varphi(x) = \begin{cases} \frac{1}{2}(x-1)^2 & \text{if } x \in [0, \infty[\\ +\infty & \text{if } x \in]-\infty, 0], \end{cases}$$

and consider the set M defined by

$$M := \left\{ Q \in \mathcal{M} \text{ such that } \int dQ = 1 \text{ and } \int (x-1/4) dQ(x) = 0 \right\}.$$

⁸ here δ is defined in the condition (3.1).

We apply the preceding results pertaining to the characterization of the projection of P on M . By Theorem 2.7, there exists a χ_+^2 -projection, say Q_+^* , of P on M . By Theorem 3.4, there exists two real numbers c_0 and c_1 such that

$$(3.12) \quad \frac{dQ_+^*}{dP}(x) \mathbb{1}_{\{q^*(x) > 0\}} = c_0 + c_1 x.$$

The support of Q_+^* is different from the support of P ; it is strictly included in $[0, 1]$. Indeed, if the support of Q_+^* is $[0, 1]$, then

$$dQ_+^*(x) = (c_0 + c_1 x) dP(x) = (c_0 + c_1 x) \mathbb{1}_{[0,1]}(x) dx.$$

Using the fact that Q_+^* belongs to M , we obtain that $c_0 = 5/2$ and $c_1 = -3$. So, Q_+^* satisfying $dQ_+^*(x) = (5/2 - 3x) dP(x)$ does not belong to $\text{dom} \chi_+^2$ (it is not a p.m.), a contradiction with the existence of the projection. This proves that the support of Q_+^* is strictly included in $[0, 1]$.

Consider now the χ^2 -divergence, i.e., the divergence associated to the convex function

$$x \in]-\infty, +\infty[\mapsto \varphi(x) = \frac{1}{2}(x-1)^2,$$

and the set M^1 defined by

$$M^1 := \left\{ Q \in \mathcal{M}^1 \text{ such that } \int dQ = 1 \text{ and } \int (x-1/4) dQ(x) = 0 \right\}.$$

Note that minimizing $\chi^2(\cdot, P)$ on M^1 is equivalent to minimizing $\chi_+^2(\cdot, P)$ on M . Hence, Q_+^* is the χ^2 -projection of P on M^1 , it has not the same support as P and (3.12) is not a definite description of the projection. On the other hand, the χ^2 -projection, say Q^* , of P on M exists, it has the same support as P , it is a signed measure and it is characterized by $dQ^*(x) = (5/2 - 3x) dP(x)$. This example shows the interest of enhancing M^1 to M .

4. FENCHEL DUALITY FOR ϕ -DIVERGENCES

We refer to Fenchel (1949), Moreau (1962), Brøndsted (1964), Rockafellar (1968), Rockafellar (1974) and Ekeland and Témam (1999) for the notion of Fenchel duality of general convex functions on general vector spaces. We consider this notion for ϕ -divergences functionals $Q \mapsto \phi(Q, P)$ viewed as convex functions on the vector space of signed finite measures $\mathcal{M}_{\mathcal{F}}$; we give different versions of dual representations of the ϕ -divergences (see Theorems 4.1, 4.3, 4.4 and 4.5 below). In view of Proposition 2.1, we identify the topological dual space of $[\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}]$ with $\langle \mathcal{F} \cup \mathcal{B}_b \rangle$ and the topological dual space of $[\langle \mathcal{F} \cup \mathcal{B}_b \rangle; \tau_{\mathcal{M}}]$ with $\mathcal{M}_{\mathcal{F}}$. Hence, the Fenchel-Legendre transform (i.e., the conjugate) of the convex function $Q \in [\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}] \mapsto \phi(Q, P) \in [0, +\infty]$ is defined as follows

$$(4.1) \quad f \in [\langle \mathcal{F} \cup \mathcal{B}_b \rangle; \tau_{\mathcal{M}}] \mapsto \phi^*(f) := \sup_{Q \in \mathcal{M}_{\mathcal{F}}} \left\{ \int f dQ - \int \varphi \left(\frac{dQ}{dP} \right) dP \right\},$$

which is convex and lower semi-continuous⁹ w.r.t. the $\tau_{\mathcal{M}}$ -topology, the weak topology induced on $\langle \mathcal{F} \cup \mathcal{B}_b \rangle$ by $\mathcal{M}_{\mathcal{F}}$.

By the lower semi-continuity of the convex function $Q \in [\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}] \mapsto \phi(Q, P) \in [0, +\infty]$ (see Proposition 2.2 above), applying the Fenchel duality theory (see e.g. Rockafellar (1968), Fenchel (1949) or Dembo and Zeitouni (1998) Lemma 4.5.8), we can state the following result for any ϕ -divergence.

⁹Note that the conjugate of a convex function is always l.s.c. w.r.t. the weak topology.

Theorem 4.1. *The function $Q \in [\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}] \mapsto \phi(Q, P) \in [0, +\infty]$ is the conjugate of its conjugate $f \in [\langle \mathcal{F} \cup \mathcal{B}_b \rangle; \tau_{\mathcal{M}}] \mapsto \phi^*(f)$ defined by (4.1). In other words, the ϕ -divergence $\phi(Q, P)$ admits the dual representation*

$$(4.2) \quad \phi(Q, P) = \sup_{f \in \langle \mathcal{F} \cup \mathcal{B}_b \rangle} \left\{ \int f \, dQ - \phi^*(f) \right\}, \text{ for all } Q \in \mathcal{M}_{\mathcal{F}},$$

where $\phi^*(\cdot)$ is defined by (4.1).

We now turn to the calculation of $\phi^*(f)$ (in particular the equality $\phi^*(f) = \int \varphi^*(f) \, dP$), and the problems of existence, uniqueness and characterization of a dual optimal solution in (4.2) (i.e., a function $\bar{f} \in \langle \mathcal{F} \cup \mathcal{B}_b \rangle$ which realizes the supremum in (4.2)) when $\phi(Q, P)$ is finite.

In the following Proposition, when φ is strictly convex and differentiable, we give the explicit form of $\phi^*(f)$ for all $f \in \langle \mathcal{F} \cup \mathcal{B}_b \rangle$ such that $\text{Im} f \subseteq \text{Im} \varphi'$.

Proposition 4.2. *Assume that φ is strictly convex and differentiable, and that for all $f, g \in \langle \mathcal{F} \cup \mathcal{B}_b \rangle$ such that $\text{Im} f \subseteq \text{Im} \varphi'$, the integrals*

$$(4.3) \quad \int |g| |\varphi'^{-1}(f)| \, dP \text{ and } \int \varphi(\varphi'^{-1}(f)) \, dP \text{ are finite.}$$

Then for all $f \in \langle \mathcal{F} \cup \mathcal{B}_b \rangle$ such that $\text{Im} f \subseteq \text{Im} \varphi'$, we have

$$(4.4) \quad \phi^*(f) \text{ is finite, and } \phi^*(f) = \int \varphi^*(f) \, dP = \int [f \varphi'^{-1}(f) - \varphi(\varphi'^{-1}(f))] \, dP.$$

Proof of Proposition 4.2 For all f in $\langle \mathcal{F} \cup \mathcal{B}_b \rangle$, define the mapping $G_f : \mathcal{M}_{\mathcal{F}} \rightarrow]-\infty, +\infty]$ by

$$G_f(Q) := \phi(Q, P) - \int f \, dQ,$$

from which $\phi^*(f) = -\inf_{Q \in \mathcal{M}_{\mathcal{F}}} G_f(Q)$. The function $G_f(\cdot)$ is strictly convex. Its domain is

$$\text{dom} G_f := \{Q \in \mathcal{M}_{\mathcal{F}} \text{ such that } G_f(Q) < +\infty\}.$$

Denote by $Q_0 := \arg \inf_{Q \in \mathcal{M}_{\mathcal{F}}} G_f(Q)$, which belongs to $\text{dom} G_f$, if it exists. It follows that Q_0 is a.c. w.r.t. P . Since $\mathcal{M}_{\mathcal{F}}$ is convex set, the measure Q_0 (if it exists) is the only measure in $\text{dom} G_f$ such that for any measure R in $\text{dom} G_f$,

$$G'_f(Q_0, R - Q_0) \geq 0,$$

where $G'_f(Q_0, R - Q_0)$ is the directional derivative of the function G_f at point Q_0 in direction $R - Q_0$; see e.g. Theorem III.31 in Azé (1997). Denote $r := \frac{dR}{dP}$ and $q_0 := \frac{dQ_0}{dP}$. By its very definition, we have

$$\begin{aligned} G'_f(Q_0, R - Q_0) &:= \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \{G_f(Q_0 + \epsilon(R - Q_0)) - G_f(Q_0)\} \\ &= \lim_{\epsilon \downarrow 0} \int \frac{1}{\epsilon} [\varphi(q_0 + \epsilon(r - q_0)) - \varphi(q_0)] \, dP - \int f \, d(R - Q_0). \end{aligned}$$

Define the function

$$g(\epsilon) := \frac{1}{-\epsilon} [\varphi(q_0 + \epsilon(r - q_0)) - \varphi(q_0)].$$

Convexity of φ implies

$$g(\epsilon) \uparrow \varphi'(q_0)(q_0 - r) \text{ when } \epsilon \downarrow 0,$$

and for all $0 < \epsilon \leq 1$ and R in $\text{dom} G_f$, we have

$$g(\epsilon) \geq g(1) = -(\varphi(r) - \varphi(q_0)) \in L_1(\mathcal{X}, P).$$

So, applying the monotone convergence theorem, we obtain

$$G'_f(Q_0, R - Q_0) = \int (\varphi'(q_0) - f) d(R - Q_0) \geq 0.$$

Therefore, under assumption (4.3), for any function f in $\langle \mathcal{F} \cup \mathcal{B}_b \rangle$ such that $\text{Im} f \subset \text{Im} \varphi'$, the measure Q_0 exists and it is given by $dQ_0 = \varphi'^{-1}(f) dP$. It follows that

$$(4.5) \quad \varphi^*(f) = \int \varphi^*(f) dP = \int \left[f \varphi'^{-1}(f) - \varphi \left(\varphi'^{-1}(f) \right) \right] dP. \quad \blacksquare$$

Remark 4.1. If the convex function $Q \in \mathcal{M}_{\mathcal{F}} \mapsto \phi(Q, P) \in [0, +\infty]$ is proper, i.e.,

$$(4.6) \quad \text{there exists at least one measure } Q_0 \text{ in } \mathcal{M}_{\mathcal{F}}(P) \text{ such that } \phi(Q_0, P) \text{ is finite,}$$

then the integral $\int \varphi^*(f) dP$ is well defined for all $f \in \langle \mathcal{F} \cup \mathcal{B}_b \rangle$. Indeed, for all $f \in \langle \mathcal{F} \cup \mathcal{B}_b \rangle$ and for all $x \in \mathcal{X}$, by Fenchel's inequality, we have

$$(4.7) \quad \varphi^*(f(x)) \geq f(x) \frac{dQ_0}{dP}(x) - \varphi \left(\frac{dQ_0}{dP}(x) \right).$$

The RHS term belong to $L_1(\mathcal{X}, P)$ by assumption (4.6). Hence, the integral $\int \varphi^*(f) dP$ is well defined. Moreover, we have $-\infty < \int \varphi^*(f) dP \leq +\infty$ for all $f \in \langle \mathcal{F} \cup \mathcal{B}_b \rangle$. Hence, from Theorem 4.1 we can state the following result.

Theorem 4.3. Assume that φ is differentiable. Then, for all $Q \in \mathcal{M}_{\mathcal{F}}$ such that $\phi(Q, P)$ is finite and $\varphi' \left(\frac{dQ}{dP} \right)$ belongs to $\langle \mathcal{F} \cup \mathcal{B}_b \rangle$, the ϕ -divergence $\phi(Q, P)$ admits the dual representation

$$(4.8) \quad \phi(Q, P) = \sup_{f \in \langle \mathcal{F} \cup \mathcal{B}_b \rangle} \left\{ \int f dQ - \int \varphi^*(f) dP \right\},$$

and the function $\bar{f} := \varphi' \left(\frac{dQ}{dP} \right)$ is a dual optimal solution. Furthermore, if φ is essentially smooth, then \bar{f} is the unique dual optimal solution (P -a.e.).

Proof of Theorem 4.3 Let $Q \in \mathcal{M}_{\mathcal{F}}$ such that $\phi(Q, P)$ is finite. Then, the integral $\int \varphi^*(f) dP$ is well defined for all $f \in \langle \mathcal{F} \cup \mathcal{B}_b \rangle$; see Remark 4.1. Furthermore, using (4.7) for all $Q \in \mathcal{M}_{\mathcal{F}}(P)$, we can see that $\phi^*(f) \leq \int \varphi^*(f) dP$ for all $f \in \langle \mathcal{F} \cup \mathcal{B}_b \rangle$. Hence, using Theorem 4.1, we can write

$$\phi(Q, P) = \sup_{f \in \langle \mathcal{F} \cup \mathcal{B}_b \rangle} \left\{ \int f dQ - \phi^*(f) \right\} \geq \sup_{f \in \langle \mathcal{F} \cup \mathcal{B}_b \rangle} \left\{ \int f dQ - \int \varphi^*(f) dP \right\}.$$

On the other hand, by (1.10), we obtain for the function $\bar{f} := \varphi' (dQ/dP)$,

$$\varphi^*(\bar{f}) = \varphi' \left(\frac{dQ}{dP} \right) \frac{dQ}{dP} - \varphi \left(\frac{dQ}{dP} \right).$$

From this, using the fact that the integrals $\int \left| \varphi' \left(\frac{dQ}{dP} \right) \right| d|Q|$ and $\int \varphi \left(\frac{dQ}{dP} \right) dP$ are finite, by simple calculus we obtain the equality $\int \bar{f} dQ - \int \varphi^*(\bar{f}) dP = \phi(Q, P)$, which completes the proof. \blacksquare

Theorem 4.3 remains valid if we substitute the vector space $\langle \mathcal{F} \cup \mathcal{B}_b \rangle$ by the arbitrary class of function \mathcal{F} . We state this result in the following Theorem.

Theorem 4.4. *Assume that φ is differentiable. Let \mathcal{F} be an arbitrary class of measurable real valued functions on \mathcal{X} . Then, for all $Q \in \mathcal{M}_{\mathcal{F}}$ such that $\phi(Q, P)$ is finite and $\varphi' \left(\frac{dQ}{dP} \right)$ belongs to \mathcal{F} , the ϕ -divergence $\phi(Q, P)$ admits the dual representation*

$$(4.9) \quad \phi(Q, P) = \sup_{f \in \mathcal{F}} \left\{ \int f \, dQ - \int \varphi^*(f) \, dP \right\},$$

and the function $\bar{f} := \varphi' \left(\frac{dQ}{dP} \right)$ is a dual optimal solution. Furthermore, if φ is essentially smooth, then \bar{f} is the unique dual optimal solution (P -a.e.).

Remark 4.2. *Theorem 4.4, with an appropriate choice of the class \mathcal{F} , has been used by Keziou (2003a) and Broniatowski and Keziou (2003) to introduce a new common definition of the “minimum ϕ -divergence estimates” in discrete or continuous parametric models. Note that the “plug-in” minimum ϕ -divergence estimates introduced by Liese and Vajda (1987) in chapter 10 are defined only in discrete parametric models, see also Lindsay (1994) and Morales et al. (1995). The use of the dual representation (4.9) allows to give a common definition of the minimum ϕ -divergence estimates in discrete or continuous parametric models.*

Remark 4.3. *Other versions of dual representations of ϕ -divergences are given in Borwein and Lewis (1991) on $L_k(\mathcal{X}, P)$ spaces, in Borwein and Lewis (1993) on compact metric spaces, and in Léonard (2001a) on Orlicz spaces. See also Rockafellar (1968) for other convex integral functionals on some “decomposable” spaces.*

Under the assumption

$$(4.10) \quad \int_{\mathcal{X}} |f| \, dP \text{ is finite for all } f \in \mathcal{F},$$

the convex function $Q \in [\mathcal{M}_{\mathcal{F}}(P); \tau_{\mathcal{F}}] \mapsto \phi(Q, P) \in [0, +\infty]$ is proper. Its Fenchel-Legendre transform is

$$(4.11) \quad f \in [\langle \mathcal{F} \cup \mathcal{B}_b \rangle; \tau_{\mathcal{M}}] \mapsto \phi^*(f) := \sup_{Q \in \mathcal{M}_{\mathcal{F}}(P)} \left\{ \int f \, dQ - \phi(Q, P) \right\} \in (-\infty, +\infty],$$

which is convex and lower semi-continuous. Following Rockafellar (1968) p. 532, let $L^* := \langle \mathcal{F} \cup \mathcal{B}_b \rangle$ and $L := \mathcal{M}_{\mathcal{F}}(P)$. Then condition (4.10) implies that both L^* and L are decomposable. Hence, we can apply the Corollary of Theorem 2 in Rockafellar (1968), to obtain the following result:

Theorem 4.5. *Under assumption (4.10), the convex conjugate function $f \in \langle \mathcal{F} \cup \mathcal{B}_b \rangle \mapsto \phi^*(f)$ defined by (4.11) is proper, it can be expressed by*

$$(4.12) \quad \phi^*(f) = \int \varphi^*(f) \, dP \text{ for all } f \in \langle \mathcal{F} \cup \mathcal{B}_b \rangle,$$

and the ϕ -divergence $\phi(Q, P)$ admits the dual representation

$$(4.13) \quad \phi(Q, P) = \sup_{f \in \langle \mathcal{F} \cup \mathcal{B}_b \rangle} \left\{ \int f \, dQ - \int \varphi^*(f) \, dP \right\}, \text{ for all } Q \in \mathcal{M}_{\mathcal{F}}(P).$$

In particular, the function $Q \in [\mathcal{M}_{\mathcal{F}}(P); \tau_{\mathcal{F}}] \mapsto \phi(Q, P) \in [0, +\infty]$ is lower semi-continuous.

Remark 4.4. *The lower semi-continuity property of the function*

$$Q \in [\mathcal{M}_{\mathcal{F}}(P); \tau_{\mathcal{F}}] \mapsto \phi(Q, P) \in [0, +\infty]$$

holds from Proposition 2.2 and Lemma 2.3 without assuming (4.10). On the other hand, Theorem 4.4 and 4.5 are of interest particularly when $\phi(Q, P)$ is finite and the class \mathcal{F} contains the function $f = \varphi'(dQ/dP)$. In Theorem 4.4, condition on Q , i.e., $\int |\varphi'(dQ/dP)| d|Q| < \infty$, holds whenever $\phi(Q, P)$ is finite and φ satisfies condition (3.1); see Lemma 3.2. However, in Theorem 4.5, these conditions do not inevitably imply assumption (4.10) if the class \mathcal{F} contains $\varphi'(dQ/dP)$. It is the case, for example, when $\phi = KL$, Q is a normal law and P is a Cauchy law. Indeed, $KL(Q, P)$ is finite, the assumption $\int |\log(dQ/dP)| dQ < \infty$ in Theorem 4.4 holds while the assumption $\int |\log(dQ/dP)| dP < \infty$ in Theorem 4.5 does not. This shows the interest of Proposition 2.2 and Theorem 4.4.

5. APPLICATIONS TO THE MINIMIZATION OF ϕ -DIVERGENCES ON SETS OF SIGNED FINITE MEASURES SATISFYING LINEAR CONSTRAINTS

In this section we apply the results of the sections 2, 3 and 4 to the optimization problem

$$\inf_{Q \in M_g} \phi(Q, P)$$

where M_g is defined in (1.13).

Under different assumptions, we obtain the dual equality $\inf(1.14) = \sup(1.15)$ and results about the problems of existence, uniqueness and characterization of the dual optimal solution and the ϕ -projections of P on the set M_g .

We state our results under the following assumptions:

- (5.1) the convex function φ is differentiable;
- (5.2) there exists at least one ϕ -projection Q^* of P on M_g with the same support as P .

Theorem 5.1. *Assume that conditions (3.1), (5.1) and (5.2) hold. Then*

- (1) *there exists $\bar{\lambda} \in \mathbb{R}^{1+l}$ such that*

$$\varphi' \left(\frac{dQ^*}{dP}(x) \right) = \bar{\lambda}_0 + \sum_{i=1}^l \bar{\lambda}_i g_i(x) \quad (P - a.e.),$$

- (2) *the equality*

$$\inf_{Q \in M_g} \phi(Q, P) = \sup_{\lambda \in \mathbb{R}^{1+l}} \left\{ \lambda_0 - \int_{\mathcal{X}} \varphi^* (\lambda^T g(x)) dP(x) \right\}$$

holds, and $\bar{\lambda}$ is a dual optimal solution. Furthermore, if the function φ is essentially smooth, then the dual optimal solution $\bar{\lambda}$ is unique.

Remark 5.1. *Under assumptions of Theorem 5.1, the ϕ -projection of P on M_g is characterized without supposing that $\bar{\lambda}$ is an interior point of $\text{dom} \phi^*$. Furthermore, the dual equality holds and the dual optimal solution is attained. Sufficient conditions for assumption (5.2) are given in Corollary 5.2 and Proposition 5.3 below.*

Proof of Theorem 5.1 Under assumptions (3.1) and (5.1), part (1) is a direct consequence of Theorem 3.4 part (2). We prove now part (2). We have $\inf_{Q \in M_g} \phi(Q, P) = \phi(Q^*, P)$ since Q^* is a ϕ -projection of P on M_g . Now, by Theorem 4.4, choosing the class of measurable functions

$$\mathcal{F} = \{x \in \mathcal{X} \mapsto \lambda^T g(x) \text{ such that } \lambda \in \mathbb{R}^{1+l}\},$$

we can write

$$\phi(Q^*, P) = \sup_{\lambda \in \mathbb{R}^{1+l}} \left\{ \lambda_0 - \int_{\mathcal{X}} \varphi^* (\lambda^T g(x)) \, dP(x) \right\},$$

and from it we deduct that $\bar{\lambda}$ is a dual optimal solution by the same Theorem. ■

Corollary 5.2. *Assume that φ is differentiable and strictly convex. If there exists some $\bar{\lambda} \in \mathbb{R}^{1+l}$ such that*

$$(5.3) \quad \int \varphi \left(\varphi'^{-1} \left(\bar{\lambda}^T g(x) \right) \right) \, dP < \infty \quad \text{and} \quad \int g^T \varphi'^{-1} \left(\bar{\lambda}^T g(x) \right) \, dP(x) = (1, 0, \dots, 0)^T,$$

then

- (1) *the measure Q^* defined by $dQ^*(x) = \varphi'^{-1} \left(\bar{\lambda}^T g(x) \right) \, dP(x)$ is the unique ϕ -projection of P on M_g .*
- (2) *the equality*

$$\inf_{Q \in M_g} \phi(Q, P) = \sup_{\lambda \in \mathbb{R}^{1+l}} \left\{ \lambda_0 - \int_{\mathcal{X}} \varphi^* (\lambda^T g(x)) \, dP(x) \right\}$$

holds, and $\bar{\lambda}$ is a dual optimal solution. Furthermore, if the function φ is essentially smooth, then the dual optimal solution $\bar{\lambda}$ is unique.

In particular, (5.3) holds if there exists a dual optimal solution $\bar{\lambda}$ which is an interior point of

$$\text{dom} \phi^* := \left\{ \lambda \in \mathbb{R}^{1+l} \text{ such that } \int_{\mathcal{X}} |\varphi^* (\lambda^T g(x))| \, dP(x) \text{ is finite} \right\}.$$

Proof of Corollary 5.2 (1) Apply Theorem 3.4 part (1). (2) the proof is the same as that of part (2) of Theorem 5.1. ■

Remark 5.2. *Note that, if φ is differentiable and strictly convex, then under assumption (3.1), conditions (5.2) and (5.3) are equivalent; see Theorem 3.4 part (1) and (2).*

In the the following Proposition we give other sufficient conditions for assumption (5.2). The conditions are

- (5.4) $\phi(M_g, P) < \infty$;
- (5.5) $\lim_{|x| \rightarrow \infty} \frac{\varphi(x)}{|x|} = +\infty$;
- (5.6) for every $\alpha > 0$, and all $i = 1, \dots, l$, $\int \varphi^*(\alpha|g_i|) dP < \infty$;
- (5.7) there exists numbers $1 < r, k < +\infty$ such that $r^{-1} + k^{-1} = 1$,
 $\lim_{|x| \rightarrow \infty} \frac{\varphi(x)}{|x|^r} > 0$, and for all $i = 1, \dots, l$, $\|g_i\|_k < \infty$;
- (5.8) the functions g_1, \dots, g_l belong to $L_\infty(\mathcal{X}, P)$;
- (5.9) $\varphi(0) = +\infty$;
- (5.10) $a_\varphi = 0$ and $\varphi'(0) = -\infty$;
- (5.11) there exists some $Q_0 \in M \cap \text{dom}\phi$ such that $\frac{dQ_0}{dP} > 0$ (P -a.e.).

Proposition 5.3.

- (1) Under assumptions (5.4), (5.5), (5.6) and (5.9), condition (5.2) holds.
- (2) Condition (5.2) holds also under assumptions (5.4), (5.7) and (5.9).
- (3) Condition (5.2) holds also if, in part (1), (5.6) is replaced by (5.8) or/and if condition (5.9) is replaced by [(5.1), (5.10) and (5.11)].
- (4) Condition (5.2) holds also if, in part (2), condition (5.9) is replaced by [(5.1), (5.10) and (5.11)].

Proof of Proposition 5.3 (1) Since M_g is closed in $[\mathcal{M}_{\mathcal{F}}; \tau_{\mathcal{F}}]$ (choosing the class $\mathcal{F} = \{g_1, \dots, g_l\}$), we can then apply Theorem 2.6 to deduce that there exists at least one ϕ -projection of P on M_g . Condition (5.9) implies that Q^* has the same support as P . (2) We can apply Theorem 2.7. (3) Under assumption (5.8), the set M_g is closed in τ -topology. Hence, we can apply Theorem 2.5 to deduce that there exists at least one ϕ -projection of P on M_g . Conditions (5.1), (5.10) and (5.11) imply that Q^* has the same support as P (see Lemma 3.7). ■

REFERENCES

- Azé, D. (1997). *Eléments d'analyse convexe et variationnelle*. Ellipses, Paris.
- Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.*, **5**(3), 445–463.
- Bertail, P. (2003). Empirical likelihood in some semiparametric models. *CREST working paper n° 2003-12*.
- Bertail, P. (2004). Empirical likelihood in nonparametric and semiparametric models. In *Parametric and semiparametric models with applications to reliability, survival analysis, and quality of life*, Stat. Ind. Technol., pages 291–306. Birkhäuser Boston, Boston, MA.
- Borwein, J. M. and Lewis, A. S. (1991). Duality relationships for entropy-like minimization problems. *SIAM J. Control Optim.*, **29**(2), 325–338.
- Borwein, J. M. and Lewis, A. S. (1992). Partially finite convex programming. II. Explicit lattice models. *Math. Programming*, **57**(1, Ser. B), 49–83.
- Borwein, J. M. and Lewis, A. S. (1993). Partially-finite programming in L_1 and the existence of maximum entropy estimates. *SIAM J. Optim.*, **3**(2), 248–267.

- Brezis, H. (1983). *Analyse fonctionnelle*. Collection Mathématiques Appliquées pour la Maîtrise. Masson, Paris. Théorie et applications.
- Brøndsted, A. (1964). Conjugate convex functions in topological vector spaces. *Mat.-Fys. Medd. Danske Vid. Selsk.*, **34**(2), 27 pp. (1964).
- Broniatowski, M. (2003). Estimation of the Kullback-Leibler divergence. *Math. Methods Statist.*, **12**(4), 391–409 (2004).
- Broniatowski, M. and Keziou, A. (2003). Parametric estimation and testing through divergences. *Preprint 2004-1, L.S.T.A - Université Paris 6*.
- Broniatowski, M. and Keziou, A. (2004). Estimation and tests for models satisfying linear constraints with unknown parameter. *Preprint 2004-2, L.S.T.A - Université Paris 6*.
- Cressie, N. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc. Ser. B*, **46**(3), 440–464.
- Csiszár, I. (1963). Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, **8**, 85–108.
- Csiszár, I. (1967a). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, **2**, 299–318.
- Csiszár, I. (1967b). On topology properties of f -divergences. *Studia Sci. Math. Hungar.*, **2**, 329–339.
- Csiszár, I. (1975). I -divergence geometry of probability distributions and minimization problems. *Ann. Probability*, **3**, 146–158.
- Csiszár, I. (1984). Sanov property, generalized I -projection and a conditional limit theorem. *Ann. Probab.*, **12**(3), 768–793.
- Csiszár, I. (1995). Generalized projections for non-negative functions. *Acta Math. Hungar.*, **68**(1–2), 161–186.
- Csiszár, I., Gamboa, F., and Gassiat, E. (1999). MEM pixel correlated solutions for generalized moment and interpolation problems. *IEEE Trans. Inform. Theory*, **45**(7), 2253–2270.
- Dembo, A. and Zeitouni, O. (1998). *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics*. Springer-Verlag, New York, second edition.
- Dunford, N. and Schwartz, J. (1962). *Linear Operators*. Interscience Publishers.
- Eichelsbacher, P. and Schmock, U. (2002). Large deviations of U -empirical measures in strong topologies and applications. *Ann. Inst. H. Poincaré Probab. Statist.*, **38**(5), 779–797.
- Ekeland, I. and Témam, R. (1999). *Convex analysis and variational problems*, volume 28 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, english edition. Translated from the French.
- Fenchel, W. (1949). On conjugate convex functions. *Canadian J. Math.*, **1**, 73–77.
- Gänssler, P. (1971). Compactness and sequential compactness in spaces of measures. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **17**, 124–146.
- Groeneboom, P., Oosterhoff, J., and Ruymgaart, F. H. (1979). Large deviation theorems for empirical probability measures. *Ann. Probab.*, **7**(4), 553–586.
- Jiménez, R. and Shao, Y. (2001). On robustness and efficiency of minimum divergence estimators. *Test*, **10**(2), 241–248.
- Keziou, A. (2003a). Dual representation of ϕ -divergences and applications. *C. R. Math. Acad. Sci. Paris*, **336**(10), 857–862.
- Keziou, A. (2003b). Utilisation des divergences entre mesures en statistique inferentielle. *Doctorat de Math. d'Univ. Paris 6. Stat.*
- Léonard, C. (2001a). Convex conjugates of integral functionals. *Acta Math. Hungar.*, **93**(4), 253–280.
- Léonard, C. (2001b). Minimization of energy functionals applied to some inverse problems. *Appl. Math. Optim.*, **44**(3), 273–297.

- Léonard, C. (2001c). Minimizers of energy functionals. *Acta Math. Hungar.*, **93**(4), 281–325.
- Liese, F. (1977). On the existence of f -projections. In *Topics in information theory (Second Colloq., Keszthely, 1975)*, pages 431–446. Colloq. Math. Soc. János Bolyai, Vol. 16. North-Holland, Amsterdam.
- Liese, F. and Vajda, I. (1987). *Convex statistical distances*, volume 95. BSB B. G. Teubner Verlagsgesellschaft, Leipzig.
- Lindsay, B. G. (1994). Efficiency versus robustness: the case for minimum Hellinger distance and related methods. *Ann. Statist.*, **22**(2), 1081–1114.
- Meyer, P.-A. (1966). *Probabilités et potentiel*. Publications de l’Institut de Mathématique de l’Université de Strasbourg, No. XIV. Actualités Scientifiques et Industrielles, No. 1318. Hermann, Paris.
- Morales, D., Pardo, L., and Vajda, I. (1995). Asymptotic divergence of estimates of discrete distributions. *J. Statist. Plann. Inference*, **48**(3), 347–369.
- Moreau, J. (1962). Fonctions convexes en dualité. *Séminaire de Mathématiques de la Faculté des Sciences de Montpellier*, **no. 1**.
- Owen, A. B. (2001). *Empirical Likelihood*. Chapman and Hall, New York.
- Rockafellar, R. T. (1968). Integrals which are convex functionals. *Pacific J. Math.*, **24**, 525–539.
- Rockafellar, R. T. (1970). *Convex analysis*. Princeton University Press, Princeton, N.J.
- Rockafellar, R. T. (1974). *Conjugate duality and optimization*. Society for Industrial and Applied Mathematics, Philadelphia, Pa. Lectures given at the Johns Hopkins University, Baltimore, Md., June, 1973, Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 16.
- Rüschendorf, L. (1984). On the minimum discrimination information theorem. *Statist. Decisions*, (suppl. 1), 263–283. Recent results in estimation theory and related topics.
- Rüschendorf, L. (1987). Projections of probability measures. *Statistics*, **18**(1), 123–129.
- Teboulle, M. and Vajda, I. (1993). Convergence of best ϕ -entropy estimates. *IEEE Trans. Inform. Theory*, **39**(1), 297–301.

*LSTA-UNIVERSITÉ PARIS 6. E-MAIL: MICHEL.BRONIATOWSKI@UPMC.FR

**LABORATOIRE DE MATHÉMATIQUES, UNIVERSITÉ DE REIMS AND LSTA-UNIVERSITÉ PARIS 6., E-MAIL: AMOR.KEZIOU@UPMC.FR